

METHOD AND DEVICE FOR SUMMARIZING INFORMATION, WEIGHTING METHOD AND CHARACTER BROADCASTING RECEIVER

Publication number: JP8329118

Publication date: 1996-12-13

Inventor: IMANAKA TAKESHI; KATAOKA MITSUTERU;
MATSUURA SATOSHI

Applicant: MATSUSHITA ELECTRIC IND CO LTD

Classification:

- international: G06F17/27; G06F17/28; G06F17/27; G06F17/28;
(IPC1-7): G06F17/30; H04N7/025; H04N7/03;
H04N7/035

- european: G06F17/27A4; G06F17/27S2

Application number: JP19950253981 19950929

Priority number(s): JP19950253981 19950929; JP19940285718 19941118;
JP19950066340 19950324

Also published as:

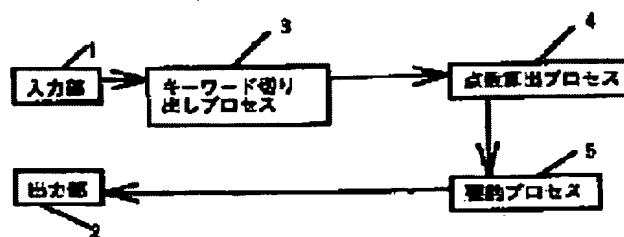


US6064952 (A1)
CN1178164C ((

Report a data error he

Abstract of JP8329118

PURPOSE: To extract and display a keyword more effectively expressing a common topic among plural units as a summary when many character string data divided into prescribed units are applied. **CONSTITUTION:** This system is provided with an input part 1 capable of receiving the input of character string data divided into prescribed units and expressing each character by a character code, an output part 2 for displaying an information summarizing result, a keyword segmenting part 3 for extracting a keyword included in each prescribed unit from character string data inputted from an input part 1, a point calculating part 4 for calculating the point of each keyword so that a keyword extracted from more units has a higher point, and a summarizing part 5 for selecting a keyword based upon a calculated point and outputting the keyword as an information summary.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-329118

(43) 公開日 平成8年(1996)12月13日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30		9194-5L	G 0 6 F 15/401	3 2 0 A
H 0 4 N 7/025		9194-5L		3 1 0 A
7/03			H 0 4 N 7/08	A
7/035				

審査請求 有 請求項の数53 O L (全 61 頁)

(21) 出願番号 特願平7-253981

(22) 出願日 平成7年(1995)9月29日

(31) 優先権主張番号 特願平6-285718

(32) 優先日 平6(1994)11月18日

(33) 優先権主張国 日本 (J P)

(31) 優先権主張番号 特願平7-66340

(32) 優先日 平7(1995)3月24日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 今中 武

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 片岡 充照

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 松浦 聡

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

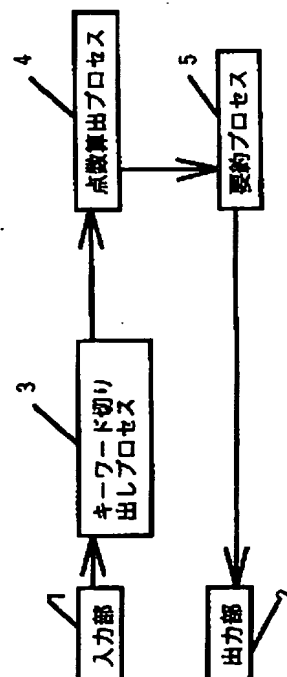
(74) 代理人 弁理士 松田 正道

(54) 【発明の名称】 情報要約方法、情報要約装置、重み付け方法、および文字放送受信装置。

(57) 【要約】

【課題】 所定の単位に区切られた文字列データが多数与えられた場合に、それらに複数の単位間で共通する話題をより効果的に表すキーワードを情報要約として抽出し、表示することができないという課題があった。

【解決手段】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力部1と、情報要約の結果を表示するための出力部2とを有し、キーワード切り出し部3において、入力部1から入力された文字列データから前記所定の単位ごとに含まれているキーワードを抽出し、点数算出部4において、より多くの単位で抽出されているキーワードほど点数が高くなるようにキーワードごとに点数を算出し、要約部5において、この算出した点数に基づいてキーワードを選択し、情報要約として出力部2に出力する構成とする。



【特許請求の範囲】

【請求項1】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付け、

その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、

その重み付けの結果に基づいてキーワードを選択し、その選択されたキーワードを前記文字列データに関する情報要約として出力することを特徴とする情報要約方法。

【請求項2】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、

前記入力手段から入力された文字列データから前記所定の単位ごとにキーワードを抽出するキーワード抽出手段と、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、

その選択されたキーワードを前記文字列データに関する情報要約として出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項3】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付け、

その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けすることを特徴とする重み付け方法。

【請求項4】 文字放送を受信する文字放送受信手段と、

所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、

前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、

前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、

その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項5】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、

前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、

又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、

それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択し、

それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として出力することを特徴とする情報要約方法。

【請求項6】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、

前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽出手段と、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、

又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

それらの重み付けの結果に基づいてキーワード、およびキーワード関連を選択する選択手段と、

それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として、出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項7】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、

前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、

前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、

又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けすることを特徴とする重み付け方法。

【請求項8】 文字放送を受信する文字放送受信手段と、

所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、

前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、

同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、

それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項9】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付け、

その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、

それら抽出したキーワード間の類似性を算出し、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、

その重み付けの結果に基づいてキーワードを選択し、

その選択されたキーワードを前記文字列データに関する情報要約として出力することを特徴とする情報要約方法。

【請求項10】 所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、

その入力された文字列データから前記所定の単位ごとにキーワードを抽出するキーワード抽出手段と、

それら抽出したキーワード間の類似性を算出する類似性算出手段と、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、

その選択されたキーワードを前記文字列データに関する情報要約として出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項11】 所定の単位に分けられており、かつ各

文字が文字コードで表された文字列データの入力を受け付け、

その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、

それら抽出したキーワード間の類似性を算出し、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けすることを特徴とする重み付け方法。

【請求項12】 文字放送を受信する文字放送受信手段と、

所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、

前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、

それら抽出したキーワード間の類似性を算出する類似性算出手段と、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、

その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項13】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、

前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、

前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出し、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、

それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択し、

それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として、出力することを特徴とする情報要約方法。

【請求項14】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、

前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽

出手段と、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出する類似性算出手段と、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、

それら選択されたキーワード、及びキーワード関連を前記文字列データに関する情報要約として出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項15】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、

前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、

前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出し、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けすることを特徴とする重み付け方法。

【請求項16】 文字放送を受信する文字放送受信手段と、

所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、

前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、

同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出する類似性算出手段と、

前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、

それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項17】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項1記載の情報要約方法。

【請求項18】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項2記載の情報要約装置。

【請求項19】 前記キーワードに重み付けする際に、その重み付けの対象となるが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項3記載の重み付け方法。

【請求項20】 前記キーワードに重み付けする際に、その重み付けの対象となるが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項4記載の文字放送受信装置。

【請求項21】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けし、

又、前記キーワード関連に重み付けする際に、その重み付けの対象となるキーワード関連が得られた前記段落の数と、前記各段落内において前記対象となるキーワード関連を構成するキーワードが出現する出現頻度と、前記対象となるキーワード関連を構成するキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項5記載の情報要約方法。

【請求項22】 前記重み付け手段は、前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けし、

又、前記キーワード関連に重み付けする際に、その重み

付けの対象となるキーワード関連が得られた前記段落の数と、前記各段落内において前記対象となるキーワード関連を構成するキーワードが出現する出現頻度と、前記対象となるキーワード関連を構成するキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項6記載の情報要約装置。

【請求項23】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けし、

又、前記キーワード関連に重み付けする際に、その重み付けの対象となるキーワード関連が得られた前記段落の数と、前記各段落内において前記対象となるキーワード関連を構成するキーワードが出現する出現頻度と、前記対象となるキーワード関連を構成するキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項7記載の重み付け方法。

【請求項24】 前記重み付け手段は、前記キーワードに重み付けする際に、その重み付けの対象となるキーワードが抽出されている前記単位の数と、前記各単位内での前記対象となるキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けし、又、前記キーワード関連に重み付けする際に、その重み付けの対象となるキーワード関連が得られた前記段落の数と、前記各段落内において前記対象となるキーワード関連を構成するキーワードが出現する出現頻度と、前記対象となるキーワード関連を構成するキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項8記載の文字放送受信装置。

【請求項25】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、その重み付けの結果に基づいてキーワード関連を選択し、その選択結果に基づいて生成した、前記文字列データに関する情報要約を出力することを特徴とする情報要約方法。

【請求項26】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段

と、前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽出手段と、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワード関連を選択する選択手段と、

その選択結果に基づいて生成された、前記文字列データに関する情報要約を出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項27】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、

前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けすることを特徴とする重み付け方法。

【請求項28】 文字放送を受信する文字放送受信手段と、

所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、

前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、

同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、

前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、

その重み付けの結果に基づいてキーワード関連を選択する選択手段と、

その選択の結果に基づいて生成された、前記文字放送に関する情報要約を表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項29】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、

同一の前記段落から得られたキーワードどうしを関連付けしてキーワード関連を生成し、

前記抽出したキーワード間の類似性を算出し、
 前記生成されたキーワード関連を構成するキーワードと、他のキーワード関連を構成するキーワードとの間の前記算出された類似性を利用して、そのキーワード関連間の類似性を算出し、
 前記キーワード関連に、そのキーワード関連と同一又は類似のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、
 その重み付けの結果に基づいてキーワード関連を選択し、
 その選択されたキーワード関連を、前記文字列データに関する情報要約として出力することを特徴とする情報要約方法。

【請求項30】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、
 前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽出手段と、
 同一の前記段落から得られたキーワードどうしを関連付けてキーワード関連を生成するキーワード関連付け手段と、
 前記抽出したキーワード間の類似性を算出する類似性算出手段と、
 前記生成されたキーワード関連を構成するキーワードと、他のキーワード関連を構成するキーワードとの間の前記算出された類似性を利用して、そのキーワード関連間の類似性を算出するキーワード関連類似性算出手段と、
 前記キーワード関連に、そのキーワード関連と同一又は類似のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、
 その重み付けの結果に基づいてキーワード関連を選択する選択手段と、
 その選択されたキーワード関連を前記文字列データに関する情報要約として出力するための出力手段と、を備えたことを特徴とする情報要約装置。

【請求項31】 所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、
 前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、
 同一の前記段落から得られたキーワードどうしを関連付けてキーワード関連を生成し、
 前記抽出したキーワード間の類似性を算出し、
 前記生成されたキーワード関連を構成するキーワードと、他のキーワード関連を構成するキーワードとの間の前記算出された類似性を利用して、そのキーワード関連間の類似性を算出し、

前記キーワード関連に、そのキーワード関連と同一又は類似のキーワード関連の他の前記段落における出現状態を考慮して、重み付けすることとを特徴とする重み付け方法。

【請求項32】 文字放送を受信する文字放送受信手段と、
 所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、
 前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、
 同一の前記段落から得られたキーワードどうしを関連付けてキーワード関連を生成するキーワード関連付け手段と、
 前記抽出したキーワード間の類似性を算出する類似性算出手段と、
 前記生成されたキーワード関連を構成するキーワードと、他のキーワード関連を構成するキーワードとの間の前記算出された類似性を利用して、そのキーワード関連間の類似性を算出するキーワード関連類似性算出手段と、
 前記キーワード関連に、そのキーワード関連と同一又は類似のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、
 その重み付けの結果に基づいてキーワード関連を選択する選択手段と、
 その選択されたキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示するための表示手段と、を備えたことを特徴とする文字放送受信装置。

【請求項33】 前記入力された文字列データを所定の作業領域に記憶し、
 前記情報要約を出力した後に、その情報要約を構成するキーワードの中から、キーワードを指定する場合、その指定されたキーワードを利用して、その指定されたキーワードを含む前記文字列データの全部又は一部を出力することを特徴とする請求項1、5、9、13、17、21、25または29記載の情報要約方法。

【請求項34】 キーワードを指定するためのキーワード指定手段と、
 前記入力手段から入力された前記文字列データを記憶するための作業領域とを有し、
 前記出力手段に前記情報要約が出力された後に、その出力された情報要約を構成するキーワードの中から、前記キーワード指定手段によって、キーワードが指定された場合、その指定されたキーワードを利用して、そのキーワードを含む前記文字列データの全部又は一部を前記出力手段によって出力することを特徴とする請求項2、6、10、14、18、22、26または30記載の情報要約装置。

【請求項35】 キーワードを指定するためのキーワー

ド指定手段と、

前記文字放送受信手段から受信された前記文字放送データを記憶するための作業領域とを有し、

前記表示手段に前記情報要約が表示された後に、その表示された情報要約を構成するキーワードの中から、前記キーワード指定手段によって、キーワードが指定された場合、その指定されたキーワードを利用して、そのキーワードを含む前記文字放送のデータの全部又は一部を前記表示手段によって表示することを特徴とする請求項4、8、12、16、20、24、28または32記載の文字放送受信装置。

【請求項36】 前記類似性のより高いキーワードで構成される前記キーワード関連が、より多くの前記段落において存在するほど、そのキーワード関連に対して、より大きく重み付けすることを特徴とする請求項13記載の情報要約方法。

【請求項37】 前記類似性のより高いキーワードで構成される前記キーワード関連が、より多くの前記段落において存在するほど、そのキーワード関連に対して、より大きく重み付けすることを特徴とする請求項14記載の情報要約装置。

【請求項38】 前記類似性のより高いキーワードで構成される前記キーワード関連が、より多くの前記段落において存在するほど、そのキーワード関連に対して、より大きく重み付けすることを特徴とする請求項15記載の重み付け方法。

【請求項39】 前記類似性のより高いキーワードで構成されるキーワード関連が、より多くの前記段落において存在するほど、そのキーワード関連に対して、より大きく重み付けすることを特徴とする請求項16記載の文字放送受信装置。

【請求項40】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードとの類似性が所定基準より高いキーワードが抽出されている単位の数と、前記対象となるキーワードとの類似性が所定基準より高いキーワードの前記各単位内での出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項9または13記載の情報要約方法。

【請求項41】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードとの類似性が所定基準より高いキーワードが抽出されている単位の数と、前記各単位内での前記対象となるキーワードとの類似性が所定基準より高いキーワードの出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項10または14記載の情報要約装置。

【請求項42】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードとの類似性が所定基準より高いキーワードが抽出されている単位の数と、

前記対象となるキーワードとの類似性が所定基準より高いキーワードの前記各単位内での出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項11または15記載の重み付け方法。

【請求項43】 前記キーワードに重み付けする際に、その重み付けの対象となるキーワードとの類似性が所定基準より高いキーワードが抽出されている単位の数と、前記対象となるキーワードとの類似性が所定基準より高いキーワードの前記各単位内での出現頻度と、前記対象となるキーワードの文字数とのうち、少なくとも1つ以上を利用して重み付けすることを特徴とする請求項12または16記載の文字放送受信装置。

【請求項44】 任意のキーワードの入力を受け付け、前記情報要約を出力した後に、前記単位の段落のデータ内容が更新され、且つその更新されたデータ内に前記入力されたキーワードが存在する場合に、そのキーワードを含む前記データの全部又は一部を出力し、もしくは、蓄積し、

又は、そのキーワードを含む前記データが存在したことを出力し、もしくは、蓄積することを特徴とする請求項1、5、9、13、17、21、25または29記載の情報要約方法。

【請求項45】 任意のキーワードを入力するキーワード入力手段と、

データを蓄積するためのデータ蓄積手段とを有し、

前記出力手段により情報要約が出力された後に、前記単位の段落のデータ内容が更新され、且つその更新されたデータ内に前記入力されたキーワードが存在する場合に、

そのキーワードを含む前記データの全部又は一部を、前記出力手段により出力し、もしくは、前記データ蓄積手段に蓄積し、

又は、そのキーワードを含む前記データが存在したことを、前記出力手段により出力し、もしくは、前記データ蓄積手段に蓄積することを特徴とする請求項2、6、10、14、18、22、26または30記載の情報要約装置。

【請求項46】 任意のキーワードを入力するキーワード入力手段と、

データを蓄積するためのデータ蓄積手段とを有し、

前記表示手段により情報要約が表示された後に、前記番組の段落のデータ内容が更新され、且つその更新されたデータ内に前記入力されたキーワードが存在する場合に、

そのキーワードを含む前記データの全部又は一部を、前記表示手段により表示し、もしくは、前記データ蓄積手段に蓄積し、

又は、そのキーワードを含む前記データが存在したこと

を、前記表示手段により表示し、もしくは、前記データ蓄積手段に蓄積することを特徴とする請求項4、8、12、16、20、24、28または32記載の文字放送受信装置。

【請求項47】 前記情報要約を出力した後に、その出力された情報要約を構成するキーワードの中からキーワードを指定し、

その後、前記単位の段落のデータ内容が更新され、且つその更新されたデータ内に前記指定されたキーワードが存在する場合に、

そのキーワードを含む前記データの全部又は一部を出力し、もしくは、

蓄積し、

又は、そのキーワードを含む前記データが存在したことを出力し、もしくは、蓄積することを特徴とする請求項1、5、9、13、17、21、25または29記載の情報要約方法。

【請求項48】 キーワードを指定するキーワード指定手段と、

データを蓄積するためのデータ蓄積手段とを有し、

前記出力手段により情報要約が出力された後に、その出力された情報要約を構成するキーワードの中から、前記キーワード指定手段によってキーワードが指定されて、その後、前記単位の段落のデータ内容が更新され、且つその更新されたデータ内に前記指定されたキーワードが存在する場合に、

そのキーワードを含む前記データの全部又は一部を、前記出力手段により出力し、もしくは、前記データ蓄積手段に蓄積し、

又は、そのキーワードを含む前記データが存在したことを、前記出力手段により出力し、もしくは、前記データ蓄積手段に蓄積することを特徴とする請求項2、6、10、14、18、22、26または30記載の情報要約装置。

【請求項49】 キーワードを指定するキーワード指定手段と、

データを蓄積するためのデータ蓄積手段とを有し、

前記表示手段により情報要約が表示された後に、その表示された情報要約を構成するキーワードの中から、前記キーワード指定手段によってキーワードが指定されて、その後、前記番組の段落のデータ内容が更新され、且つその更新されたデータ内に前記指定されたキーワードが存在する場合に、

そのキーワードを含む前記データの全部又は一部を、前記表示手段により表示し、もしくは、前記データ蓄積手段に蓄積し、

又は、そのキーワードを含む前記データが存在したことを、前記表示手段により表示し、もしくは、前記データ蓄積手段に蓄積することを特徴とする請求項4、8、12、16、20、24、28または32記載の文字放送

受信装置。

【請求項50】 前記情報要約として互いに関連する複数のキーワードを実質的に一固まりに配置し、その配置された状態で前記出力することを特徴とする請求項5、13、25または29記載の情報要約方法。

【請求項51】 前記情報要約として互いに関連する複数のキーワードが実質的に一固まりに配置され、その配置された状態で前記出力手段により出力することを特徴とする請求項6、14、26または30記載の情報要約装置。

【請求項52】 前記情報要約として互いに関連する複数のキーワードを表示部を用いて出力する場合、それら複数のキーワードのうち、1つ目のキーワードの表示位置を基準として、残りのキーワードを前記基準となる表示位置より所定の文字数だけ字下げした位置に改行して表示し、前記互いに関連する複数のキーワードを視覚的に一固まりに表示することを特徴とする請求項5、13、25または29記載の情報要約方法。

【請求項53】 前記出力手段は、前記情報要約として互いに関連する複数のキーワードを表示する表示部を有し、

その表示部に表示される前記複数のキーワードのうち、1つ目のキーワードの表示位置を基準として、残りのキーワードを前記基準となる表示位置より所定の文字数だけ字下げした位置に改行して表示し、前記互いに関連する複数のキーワードを視覚的に一固まりに表示するための表示制御部を備えたことを特徴とする請求項6、14、26または30記載の情報要約装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、例えば、文字放送受信装置など所定の単位に分けられた複数の文字列データから所定のキーワードを抽出する場合等に利用可能な、情報要約方法、情報要約装置、重み付け方法、および文字放送受信装置に関する。

【0002】

【従来の技術】近年、マルチメディア時代の到来によりCD-ROMなどのパッケージメディアのみでなく、数多くの情報が通信や放送などを通じて提供されてきている。これらの情報の中には、映像、音声に関するものに加えて電子ブック、文字放送など文字情報がある。文字情報はASCIIコード、JISコードなど計算機で容易に処理できる文字コードで表現されている。しかしながら、人間にとっては文字情報は一度に表示できる情報量が少ない、映像情報と比べて要点を解釈するのに必要な時間が多い、といった問題がある。この問題は、情報化社会が進展する中、提供される情報はますます増えて行くことを考えれば、さらに深刻になる。このような問題に対し、自動的に文書内容を解釈して分かりやすく表現し直す技術への取り組みが考えられる。このような取り組み

の例として、人工知能研究分野における自然言語処理の研究がある。しかしながら、大規模な辞書情報、文法情報などが必要である上に文書内容を誤って解釈する確率を実用レベルにまで引き下げるのが困難など現実的問題に適用するには問題点が多く、実例的な応用例は少ない。

【0003】一方、近年では文字コードが放送として送られてくる文字放送を対象とした受信装置が開発、販売されており、家庭内に向けて提供される文字情報は急激に増えてきている。文字放送では、多数の番組が提供されている反面、情報の形態が文字なので、利用者はテレビに映し出される文章を読むことによって情報を得ることができる。この結果、情報の全容を把握するためには大量の文字を読まなければならない、ページをめくりながら順に読まなければならない、などの問題が生じる。実際に、ニュースなどでは何が情報として提供されているのか予め解っていることは希であり、利用者自身があらかじめ自分の興味のある情報を意識している場合が少ない。このため、自分にとって必要な情報のみを取り出すことは困難であり、全部の情報を一通り読み終えてから、必要な情報を利用者自身が選び出すことになる。したがって、必要な情報を得るために費やす時間が長くなり、文字放送を楽しむ利用者の数が多くなることに対する妨げになっている。このため、文字放送において情報のポイントを抽出し、ポイントのみを表示するなどの情報要約機能の必要性は高まっている。また、文字放送中には主なニュースの要約を放送しているチャンネルもあるが、要約自体が数ページに渡っている、放送局によって要約の内容、形式、長さが異なるなど問題が多い。

【0004】従来より、文書データの情報要約技術として実用化されているものとしてキーワード抽出技術がある。この技術は、学術論文などを対象とし、論文文中で用いられている専門用語などの頻度を算出し、頻度の高いキーワードを論文の要約として抽出するものである。このような技術が実用化された背景には、特定分野の論文といったように専門用語がある程度限られる文書を対象としていたことがある。このような対象ではキーワードとして抽出すべき用語の辞書を予め準備することが比較的容易である。この技術により自動抽出したキーワードは各論文に付与し、論文の整理、検索に用いられる。

【0005】

【発明が解決しようとする課題】しかしながら、上述のキーワード抽出技術を文字放送番組の要約に適用した場合、番組内に現れるキーワードのうち単純に頻度の大きいものを抽出してしまう。この結果、互いに類似した内容のキーワードがいくつも抽出され、要約としては冗長になる。さらに、文字放送を対象とした場合には、番組の内容の要約以外に複数の番組に共通する話題を、その時々の話題性のある情報として抽出したいといった要求が生じる。たとえば、ニュース番組が複数のチャンネル

で放送されており、複数チャンネルのニュースから共通している話題を現在のトレンドとして抽出するような情報要約に対する要求がある。この場合、同一番組内で繰り返し登場するキーワードと複数の番組にまたがって登場するキーワードを区別して扱う技術が必要である。また、従来の自然言語処理技術を例えばニュース番組に適用しようとしても、ニュース番組には固有名詞が非常に多く用いられるといった特徴があり、予め用語の辞書を準備することはできない。この結果、従来技術はそのままでは適用できない。したがって、情報のトレンドを扱うことができ、かつ用語の辞書を必要としない情報要約技術が必要である。

【0006】また、情報要約としてキーワードを用いる場合、単純に頻度に基づくキーワードの抽出を行えば、類似の内容を示すキーワードがいくつも抽出されたり、キーワード間の関連がわからないなど、キーワードの数の割に情報として得られる内容が乏しくなるといった問題を引き起こす。たとえば、実際に同じ時間に放送されていた文字放送の7個のニュースから頻度の高いキーワードを単純に取り出した場合、頻度の高い上位7個のキーワードは図15に示すようになった。図15の(a)、(b)は異なる時間に受信したニュースに対して行った実験結果である。この結果からも単純に頻度に基づくキーワードの抽出では情報要約として問題があることが明白である。たとえば、(a)においては1番目のキーワード「泉佐野」と6番目のキーワード「朝日新聞社」は同じ話題に関連するキーワードであるが、そのことは図15のようなキーワードの羅列では解らない。このため、類似した意味を持つキーワードを重複して抽出しないようにする、キーワード間の関連を情報要約の中で明示的に示す技術が必要である。この必要性に対し、従来の自然言語処理のようにキーワードの意味や関連を記述した辞書情報を予め用意することが考えられるが、現実的な問題を考えた場合、大規模な辞書情報を準備することはコスト面からも問題になる場合が多く、可能な限り予め準備する辞書情報を少なくする必要がある。さらに、文字放送のニュース番組などを対象とする場合、固有名詞が多数用いられるために予め辞書を準備すること自体困難である。したがって、辞書を用いずにキーワード間の関連を自動的に推論し、キーワード間の関連も考慮した形で情報要約を行う技術が必要である。

【0007】さらに、以上のような技術は文字放送に限定されず、文字コードで与えられた文字列データ一般に必要である。たとえば、論文データにおいても個々の論文の要約ではなく、学会全体の話題の要約が必要になる場合もある。また、通信を用いた電子メール装置においては、最近の話題など全てのメールにおいて何が話題として多く取り上げられているのかを抽出する際にも必要である。

【0008】また、辞書を用いずにキーワードの頻度

を扱う場合に、予め類語辞書を用意しておくことはできないために、同じことを表している表現の異なるキーワードの処理が必要になる。たとえば、同じ選手の名前でも「伊達公子」と表現したり、「伊達選手」と表現したりする。このように表現が異なっている場合でも同じ意味で用いられているキーワードは、頻度を数えるときに互いの頻度を加算する必要がある。したがって、辞書を用いずに「伊達公子」と「伊達選手」を類似キーワードとして扱い、類似キーワードの頻度を考慮して頻度計算を行う技術の開発が課題である。

【0009】表現の異なるキーワードの使用は、文字放送など複数の放送局からデータが提供され、放送局が違えばデータを作成した作成者が違うといった場合には頻繁に生じる。特に、放送内容がニュースなど世の中の出来事を扱っている場合、学問分野などと異なり専門用語がきちんと決まっていけないという問題がある。したがって、同じ出来事を表現するためのキーワードであっても、情報源によって様々な表現が用いられる。この結果、文字放送を対象とした場合、表現の違うキーワードの処理を行う技術の開発は重要な課題である。

【0010】また、辞書を用いずに表現の異なるキーワードの類似性を算出する場合、2つのキーワードの間で同じ文字の数や割合を用いる方法が考えられる。たとえば、「伊達公子」と「伊達選手」は同じ文字が「伊達」の2文字であり、キーワード4文字中2文字が一致している。すなわち、半分以上の文字が一致しているために、これらのキーワードは類似していると見なすことができる。しかしながら、このように単純に一致している文字数に基づいて類似度を算出すれば、「伊達」の2文字が一致するキーワード「伊達公子」と「伊達選手」と「伊達政宗」はすべて類似度合いが同じになってしまうといった問題がある。「伊達公子」と「伊達選手」の類似度は大きく、「伊達公子」と「伊達政宗」は類似度が小さくなるような算出方法を考案することが課題である。

【0011】表現の異なるキーワードの処理は、文字放送など複数の放送局からデータが提供され、データを作成した作成者が異なる場合に、必要性が大きい。特に、放送内容がニュースなど世の中の出来事を扱っている場合、学問分野などと異なり専門用語がきちんと決まっていけないという問題がある。さらに、会議名、人名、会社名などの場合、同じようなキーワードで、全く関係のない話題で扱われていることがある。したがって、文字放送を対象とした場合、表現の違うキーワードの処理を行う技術の開発が課題である。

【0012】さらに、キーワードをデータ中から抽出し、抽出したキーワードの頻度に基づいて要約を行う際には、たとえば英語の{a, the}などの冠詞、前置詞などは頻出のキーワードになる。したがって、このような頻出であり話題を表す上で意味のないキーワードを取り

除く必要がある。

【0013】また、文字放送においても、英語の引用文が提供される場合が考えられる。その場合には、前置詞、冠詞など話題を表す上で重要でないキーワードの頻度が大きくなる恐れがあり、このようなキーワードが情報要約として出力されないようにする技術の開発が課題である。

【0014】本発明は、従来のこのような問題を考慮し、従来に比べてより一層適切なキーワードをデータ内から抽出することに利用可能な、情報要約方法、情報要約装置、重み付け方法、および文字放送受信装置を提供することを目的とする。

【0015】

【課題を解決するための手段】第1の発明は、所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、前記入力手段から入力された文字列データから前記所定の単位ごとにキーワードを抽出するキーワード抽出手段と、前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けする重み付け手段と、その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、その選択されたキーワードを前記文字列データに関する情報要約として出力するための出力手段と、を備えた情報要約装置である。

【0016】第2の発明は、文字放送を受信する文字放送受信手段と、所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けする重み付け手段と、その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示するための表示手段と、を備えた文字放送受信装置である。

【0017】第3の発明は、所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽出手段と、同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、それらの重み付けの結果に基づい

てキーワード、およびキーワード関連を選択する選択手段と、それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として、出力するための出力手段と、を備えた情報要約装置である。

【0018】第4の発明は、文字放送を受信する文字放送受信手段と、所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示するための表示手段と、を備えた文字放送受信装置である。

【0019】第5の発明は、所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力手段と、その入力された文字列データから前記所定の単位ごとにキーワードを抽出するキーワード抽出手段と、それら抽出したキーワード間の類似性を算出する類似性算出手段と、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けする重み付け手段と、その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、その選択されたキーワードを前記文字列データに関する情報要約として出力するための出力手段と、を備えた情報要約装置である。

【0020】第6の発明は、文字放送を受信する文字放送受信手段と、所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、それら抽出したキーワード間の類似性を算出する類似性算出手段と、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けする重み付け手段と、その重み付けの結果に基づいてキーワードを選択するキーワード選択手段と、その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示するための表示手段と、を備えた文字放送受信装置である。

【0021】第7の発明は、所定の単位に分けられた、かつ各单位内が所定の段落に分けられた、かつ各文字が

文字コードで表された文字列データの入力を受け付ける入力手段と、前記入力手段から入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出するキーワード抽出手段と、同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出する類似性算出手段と、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、それら選択されたキーワード、及びキーワード関連を前記文字列データに関する情報要約として出力するための出力手段と、を備えた情報要約装置である。

【0022】第8の発明は、文字放送を受信する文字放送受信手段と、所定の番組のチャンネルを複数個記憶するチャンネル記憶手段と、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出するキーワード抽出手段と、同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成するキーワード関連付け手段と、前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出する類似性算出手段と、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けする重み付け手段と、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択する選択手段と、それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示するための表示手段と、を備えた文字放送受信装置である。

【0023】第9の発明は、キーワードとして処理しないキーワードを予め記憶しておく例外キーワード記憶手段を有し、入力手段から入力された文字列データから所定の単位ごとに含まれているキーワードを抽出する際に、前記例外キーワード記憶手段に記憶されているキーワードと同一のキーワードは、抽出対象から外す情報要約装置である。

【0024】第10の発明は、キーワードとして処理しないキーワードを予め記憶しておく例外キーワード記憶手段を有し、チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとに含まれているキーワードを抽出する際

に、前記例外キーワード記憶手段に記憶されているキーワードと同一のキーワードは、抽出対象から外す文字放送受信装置である。

【0025】第1の発明によれば、例えば、所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付け、その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、その重み付けの結果に基づいてキーワードを選択し、その選択されたキーワードを前記文字列データに関する情報要約として出力する。

【0026】この結果、例えば、多くの単位にまたがって用いられているキーワードが優先的に情報要約として選出される。すなわち、多くの単位で共通の話題として現れているキーワードが選出され、文字列データ全体の傾向を表す要約情報として取り出される。

【0027】第2の発明によれば、文字放送受信手段が文字放送を受信し、チャンネル記憶手段が、所定の番組のチャンネルを複数個記憶し、キーワード抽出手段が、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出し、重み付け手段が、前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けし、キーワード選択手段が、その重み付けの結果に基づいてキーワードを選択し、表示手段が、その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示する。

【0028】これにより、例えば、より多くの番組にまたがって現れているキーワードほど点数が高くなるようにキーワードごとに点数を算出する。この結果、例えば、チャンネル記憶手段に記憶されているチャンネルの文字放送のうち多くの番組で共通して取り上げられている話題のキーワードを情報要約として取り出すことができる。すなわち、文字放送で放送される番組のように时时刻刻と変化するものに対して、最新情報のトレンドを取りだし、情報要約として表示する。

【0029】第3の発明によれば、所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、前記キーワードに、そのキーワードと同一のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、それらの重み付けの結果に基づいて、キーワード、およびキーワード

関連を選択し、それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として出力する。

【0030】これにより、例えば、より多くの単位にまたがって現れているキーワード、およびより多くの段落にまたがって現れているキーワードどうしの関連付けほど点数が高くなるように、キーワード、およびキーワードどうしの関連付けごとに点数を算出する。この算出した点数に基づいてキーワード、およびキーワードどうしの関連付けを選択し、情報要約として抽出する。この結果、例えば、出現頻度の大きいキーワードを単純に選択した場合に比べ、頻繁に関連づけて用いられる関連の深いキーワードどうしを取り出すことができ、情報要約の表示の際には関連の深いキーワードが多数ある場合に代表的なもののみを表示し、関連のあるキーワードを複数表示する時には、それらを関連づけて表示することにより情報要約の内容が冗長になったり、わかりにくくなるのを防ぐ。

【0031】第4の発明によれば、文字放送受信手段が、文字放送を受信し、チャンネル記憶手段が、所定の番組のチャンネルを複数個記憶し、キーワード抽出手段が、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出し、キーワード関連付け手段が、同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、重み付け手段が、前記キーワードに、そのキーワードと同一のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、選択手段が、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択し、表示手段が、それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示する。

【0032】これにより、例えば、より多くの番組にまたがって現れているキーワード、およびより多くの段落にまたがって現れるキーワードどうしの関連付けほど点数が高くなるように、キーワード、およびキーワードどうしの関連付けごとに点数を算出する。この算出した点数に基づいてキーワード、およびキーワードどうしの関連付けを選択し、情報要約として抽出する。この結果、例えば、出現頻度の大きいキーワードを単純に選択した場合に比べ、頻繁に同時に用いられる関連の深いキーワードを関連付けて取り出すことができる。特に、文字放送では各チャンネルでニュースなど同じ情報源から得た情報が放送され、かつ1つの事件に対して各番組とも共通して用いるキーワードが複数個あるために、キーワード間の関連付けに基づいて冗長なキーワードを要約内容から取り除くこと、関連を明らかにすることにより情報

量の多い要約を表示することができる。これにより、膨大な量の情報のポイントを短時間で把握することができる。

【0033】第5の発明によれば、所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付け、その入力された文字列データから前記所定の単位ごとにキーワードを抽出し、それら抽出したキーワード間の類似性を算出し、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、その重み付けの結果に基づいてキーワードを選択し、その選択されたキーワードを前記文字列データに関する情報要約として出力する。

【0034】これにより、例えば、完全には一致していないキーワードどうしであっても類似キーワードとして扱われ、類似キーワードが数多く抽出されていれば点数が高くなり、情報要約として表示される。

【0035】第6の発明によれば、文字放送受信手段が、文字放送を受信し、チャンネル記憶手段が、所定の番組のチャンネルを複数個記憶し、キーワード抽出手段が、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出し、類似性算出手段が、それら抽出したキーワード間の類似性を算出し、重み付け手段が、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けし、キーワード選択手段が、その重み付けの結果に基づいてキーワードを選択し、表示手段が、その選択されたキーワードの全部又は一部を前記文字放送に関する情報要約として表示する。

【0036】これにより、例えば、番組ごとに異なる表現で記述されているキーワードであっても類似しているものは頻度が加算される。

【0037】第7の発明によれば、所定の単位に分けられた、かつ各単位内が所定の段落に分けられた、かつ各文字が文字コードで表された文字列データの入力を受け付け、前記入力された文字列データから前記所定の単位の段落ごとにキーワードを抽出し、同一の前記段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出し、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記単位における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択し、それら選択されたキーワード及びキーワード関連を、前記文字列データに関する情報要約として、出力する。

【0038】これにより、例えば、番組ごとに異なる表現で記述されているキーワードであっても類似しているものが抽出されていれば点数が加算される。この加算により、情報要約を表す上で重要なキーワードが異なる表現で記述されていても高い点数が与えられる。

【0039】第8の発明によれば、文字放送受信手段が、文字放送を受信し、チャンネル記憶手段が、所定の番組のチャンネルを複数個記憶し、キーワード抽出手段が、前記チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとにキーワードを抽出し、キーワード関連付け手段が、同一番組中の同一段落から得られたキーワードどうしを関連付けして、キーワード関連を生成し、類似性算出手段が、前記抽出したキーワード間の類似性を、キーワードどうしの前記関連付けを含む複数の要因に基づいて算出し、重み付け手段が、前記キーワードに、そのキーワードと同一又は類似のキーワードの他の前記番組における出現状態を考慮して、重み付けし、又、前記キーワード関連に、そのキーワード関連と同一のキーワード関連の他の前記段落における出現状態を考慮して、重み付けし、選択手段が、それらの重み付けの結果に基づいて、キーワード、およびキーワード関連を選択し、表示手段が、それら選択されたキーワード及びキーワード関連の全部又は一部を、前記文字放送に関する情報要約として表示する。

【0040】これにより、例えば、番組ごとに異なる表現で記述されているキーワードであっても、関連付けられたキーワードを用いて類似度を算出し、類似しているものが抽出されていれば点数が加算される。特に、表現が類似していても全く異なる話題で用いられている場合には関連付けられているキーワードが類似しないために算出される類似度は小さくなり、より正確に類似度を算出することができる。

【0041】第9の発明によれば、入力手段から入力された文字列データから所定の単位ごとに含まれているキーワードを抽出する際に、前記例外キーワード記憶手段に記憶されているキーワードと同一のキーワードは、抽出対象から外す。

【0042】この結果、例えば、英語の入力データが与えられた場合に、冠詞、前置詞などを例外キーワード記憶手段に記憶させることにより、情報要約として表示されるキーワードの中に、これらのキーワードが含まれないようにできる。

【0043】第10の発明によれば、チャンネル記憶手段に記憶されている各チャンネルの番組を前記文字放送受信手段で受信し、受信した番組ごとに含まれているキーワードを抽出する際に、前記例外キーワード記憶手段に記憶されているキーワードと同一のキーワードは、抽出対象から外す。

【0044】この結果、例えば、文字放送に英語の文な

どが含まれている場合にでも、冠詞、前置詞など話題を表す上であまり意味を持たないキーワードが情報要約として用いられることがなくなる。

【0045】

【発明の実施の形態】以下、本発明の実施の形態について、図を用いて説明する。

【0046】図1は、第1の発明にかかる実施の一形態の情報要約装置のシステム構成図であり、同図を用いて、第1の発明にかかる実施の一形態としての情報要約方法、および情報要約装置について述べる。

【0047】図1において、1は、所定の単位に分けられており、かつ各文字が文字コードで表された文字列データの入力を受け付ける入力部、2は情報要約の結果を表示するための出力部、3は入力部1から入力された文字列データから前記所定の単位ごとに含まれているキーワードを抽出するキーワード切り出し部、4はより多くの単位で抽出されたキーワードほど点数が高くなるようにキーワードごとに点数を算出する点数算出部、5は点数算出部4で算出された点数に基づいてキーワードを選択し、情報要約として出力部2に出力する要約部である。

【0048】以上のように構成されたシステムが実行されるハードウェア構成を図2に示す。図2は、基本的に汎用の計算機システムの構成と同じであり、図1で示したシステム構成の構成部分として説明した入力部1と、出力部2とを備えている。図2の構成部分のうち図1のシステム構成と同一構成部分については同一番号を付しており、説明を省略する。図2において、11は出力部2に表示されるデータを記憶するVRAM、12は処理のためのプログラムやデータを実行時に記憶する主記憶装置、13はプログラムやデータを蓄積しておく外部記憶装置、14は外部記憶装置13に記憶されているプログラムを主記憶装置12に転送し実行するCPUである。

【0049】以上のように構成された情報要約方法、および情報要約装置の動作を図3のフローチャートを参照しながら説明する。

【0050】ステップa1では、入力部1を通じて文字列データが入力される。入力される文字列データは、図4に示すように所定の単位に区切られたデータである。図4では、所定の単位に区切られた文字列データがn個ある。たとえば、スポーツ新聞の内容が新聞社ごとに単位1～単位nまでに格納されて入力されるなどの例が考えられる。別の例としては、特定の分野の研究論文が学会ごとに単位1～単位nに格納して入力されるなどの例が考えられる。前者の例では、本情報要約方法、および情報要約装置によりスポーツ界の主な情報が要約情報として得られ、後者の例では個々の学会の枠を超え、特定の分野における最先端の話題は何かを情報要約の結果として得ることができる。以降のステップの説明において例題と

して用いるために、ステップa1では最近のニュースが新聞社ごとに単位1～単位nに格納されたデータが入力されたものとする。入力データの例を図5に示す。これは、現在放送されている文字放送のデータに基づいて作成したものである。

【0051】ステップa2では、ステップa1で入力された文字列データから1単位を取り出す。文字列データから1単位を取り出す際には、たとえば入力された順など予め定めた順序で一つずつ取り出す。このため、ステップa2を繰り返し実行することにより、入力データから取り出すべき単位がなくなってしまう。この場合には、フローチャートに示すようにステップa5へ進む。

【0052】ステップa3では、キーワード切り出し部3においてステップa2で取り出された1単位の文字列データからキーワードの切り出しを行う。キーワードの切り出しに関しては、従来より自然言語処理技術で種々の方法が検討されており、いずれの方法でも構わない。しかしながら、辞書情報など大規模な記憶容量を必要とするものは実現の際に大きな障害になる点、ニュースなどを対象とした場合には固有名詞が多くあらかじめ辞書情報を準備できない点、の2点を考慮し、本実施の一形態では辞書情報を用いないで字種の違いに基づいてキーワードを切り出す方法を用いている。具体的には、漢字、数字あるいはカタカナが3文字以上続いているものを切り出している。ただし、切り出した時に数字で始まるキーワードになる場合は、キーワードとして扱う対象から除外する。たとえば、ステップa2で取り出されたデータが図4に示すデータの単位1であった場合、「米朝高官協議合意」、「北朝鮮」、「核問題」、「米朝高官会議」などがキーワードとして切り出される。このように、字種の違いに基づいてキーワードを切り出した場合、データによっては無意味なキーワードが切り出されるが、無意味なキーワードの一部は後述のステップa7で、他の単位のキーワードと部分照合することによって修正される。また、ステップa3では既に切り出したキーワードと同じキーワードであっても繰り返し切り出して記憶する。

【0053】ステップa4では、点数算出部4においてステップa3で切り出したキーワードで同一のものをまとめて、その数を出現頻度として算出する。たとえば、ステップa3においてキーワード「国際平和維持軍」が4個切り出された場合、キーワード「国際平和維持軍」を1つにまとめて出現頻度を4にする。ただし、本実施の一形態では同一ではない2つのキーワードA、Bが抽出されていて、キーワードBがキーワードAの部分になっており、かつその部分がキーワードAの文字数の5割より長い場合、キーワードBの出現頻度にキーワードAの出現頻度を加えてキーワードAを消去する。たとえば、キーワード「米朝高官協議合意」が5個、キーワード「米朝高官協議」が3個抽出されている場合には、キーワード

「米朝高官協議」の出現頻度を8個にし、キーワード「米朝高官協議合意」を消去する。たとえば、ステップa3で切り出されたキーワードが「泉佐野」、「ヘリ墜落」、「大阪府警」、「泉佐野市」、「朝日新聞社取材用」、「ヘリコプター」、「新聞大会開」、「第47回新聞大会」、「京都市」、「新聞販売」、「正常化」、「特別宣言」、「電動茶」、「京都府茶協同組合」、*

*「電動茶」、「米朝高官協議合意」、「北朝鮮」、「核問題」、「米朝高官会議」、「ジュネーブ」、「調印式」になっている場合を考える。この場合、同一のキーワードをまとめて、以下の(表1)のようなキーワードカウント結果を得る。

【0054】

【表1】

キーワード	出現頻度
泉佐野	2
ヘリ墜落	1
大阪府警	1
朝日新聞社取材用	1
ヘリコプター	1
新聞大会開	1
第47回新聞大会	1
京都市	1
新聞販売	1
正常化	1
特別宣言	1
電動茶	2
京都府茶協同組合	1
米朝高官協議合意	1
北朝鮮	1
核問題	1
米朝高官会議	1
ジュネーブ	1
調印式	1

【0055】また、得られたキーワードカウント結果はFILO(First In Last Out)のスタックに記憶(プッシュ)する。

【0056】ステップa5では、ステップa4で記憶したスタックからキーワードと出現頻度の組を一つ取り出す。このスタックでは取り出したデータは消去する。このため、ステップa5を繰り返し実行するうちに取り出すべきデータがなくなる。取り出すべきデータがなくなった場合にはステップa10へ進む。

【0057】ステップa6では、ステップa5で取り出した組の項目「キーワード」の値を取り出す。

【0058】ステップa7～ステップa9は点数算出部4において実行される。ステップa7では、ステップa6で取り出したキーワードを用いてキーワード頻度表に同一のキーワードが存在するか否かを判定する。キーワード頻度表は以下の(表2)に示すような形式になっており初期値は空である。したがって、初めてこのステップが呼び出

50 される時には、必ずステップa9へ進む。

【0059】また、(表2)は図5のデータ例の単位1から切り出されたキーワードのみからなっているように見えるが、実際にはキーワード頻度表は、最終的に全ての単位から切り出されたキーワードが含まれる表になる。本実施の一形態の説明ではキーワード頻度表が大き*

*くなりすぎるために、今後もキーワード頻度表の一部を例として示す。

【0060】

【表2】

キーワード	単位内頻度	単位間共有度
泉佐野	2	0
ヘリ墜落	1	0
大阪府警	1	0
朝日新聞社取材用	1	0
ヘリコプター	1	0
新聞大会開	1	0
第47回新聞大会	1	0
京都市	1	0
新聞販売	1	0
正常化	1	0
特別宣言	1	0
電動茶	2	0
京都府茶協同組合	1	0
米朝高官協議合意	1	0
北朝鮮	1	0
核問題	1	0
米朝高官会議	1	0
ジュネーブ	1	0
調印式	1	0

【0061】また、ステップa7で行う判定方法は、ステップa6で取り出したキーワードとキーワード頻度表のキーワードの項目“キーワード”と順に比較して実行する。ただし、比較の際にはステップa3の説明で述べたようにキーワード切り出しによっては無意味なキーワードが抽出されているために、「容疑者浮」と「容疑者」のように片方の比較対象がもう一方の一部分になっており、かつ部分になっている文字数がキーワード長の5割より長い場合には一致したものと見なし、キーワードは共通している部分「容疑者」にして扱う。

【0062】ステップa8は、ステップa6で取り出したキ

40 ーワードがキーワード頻度表の中に既に登録されている場合に実行される。これは、複数の単位の中に同一のキーワードが存在した場合である。この場合には、(表2)のキーワード表形式の単位間共有度の値に1を加える。また、単位内頻度の値はステップa5で取り出した組の項目「出現頻度」の値を取り出し、単位内頻度の項目に加える。例えば、(表2)のようなキーワード頻度表に対して、ステップa5で取り出した組が(米朝高官協議、3)である場合を考える。この場合、ステップa8を実行することによって以下の(表3)のようなキーワード頻度表が得られる。

【0063】

* * 【表3】

キーワード	単位内頻度	単位間共有度
泉佐野	2	0
ヘリ墜落	1	0
大阪府警	1	0
朝日新聞社取材用	1	0
ヘリコプター	1	0
新聞大会開	1	0
第47回新聞大会	1	0
京都市	1	0
新聞販売	1	0
正常化	1	0
特別宣言	1	0
電動茶	2	0
京都府茶協同組合	1	0
米朝高官協議	4	1
北朝鮮	1	0
核問題	1	0
米朝高官会議	1	0
ジュネーブ	1	0
刷印式	1	0

【0064】（表3）中でステップa8の実行によって変化した部分は、キーワード「米朝高官協議」の行のキーワード、単位内頻度、単位間共有度である。（表2）ではキーワードが、「米朝高官協議合意」であったのに対して、ステップa7での比較の際に、「米朝高官協議」を部分的に含んでおり、その部分がキーワード長の5割よりも大きいために同じキーワードとして扱われ、共通している部分「米朝高官協議」にまとめられている。

【0065】ステップa9では、キーワード頻度表の中にステップa6で取りだしたキーワードと同一のキーワードが存在しない場合であり、この場合にはステップa6で取りだしたキーワードをキーワード頻度表に追加する。追

加する際には、単位内頻度はステップa5で取りだした組の出現頻度に設定する。また、単位間共有度は0にする。

【0066】ステップa10では、要約部5においてキーワード頻度表から単位間共有度の大きいものから順にキーワードを取り出し、表示可能な数のキーワードを出力部2に表示する。例えば、表示可能なキーワード数が3個である場合に、（表4）のようなキーワード頻度表からはキーワード「米朝高官協議」、「北朝鮮」、「核問題」が順に選択される。

【0067】

【表4】

キーワード	単位内頻度	単位間共有度
泉佐野	3	1
ヘリ墜落	1	0
大阪府警	1	0
朝日新聞社	2	1
ヘリコプター	1	0
新聞大会開	1	0
第47回新聞大会	1	0
京都市	1	0
新聞販売	1	0
正常化	2	1
特別宣言	1	0
電動茶	2	0
京都府茶協同組合	1	0
米朝高官協議	3	3
北朝鮮	3	2
核問題	3	2
米朝高官会議	1	0
ジュネーブ	2	1
調印式	1	0

【0068】また、単位間共有度が同じキーワード間では、単位内頻度の大きいものから選択する。このキーワードの選択法としては、単位間共有度と単位内頻度の重み付き和で優先度合を算出し、順番に表示する方法も考えられる。この方法では、予め重み付けを表す定数 S, T を定め、各キーワードごとに（単位内頻度 $\times S$ + 単位間共有度 $\times T$ ）を算出し、この値の大きいものから選択する。さらに、別の方法として、キーワードの長さを考慮する方法もある。この方法は、キーワードの文字数を導入した重み付き和で優先度合を算出する。たとえば、 S, T に加えて定数 U を予め定めて、（単位内頻度 $\times S$ + 単位間共有度 $\times T$ + キーワードの文字数 $\times U$ ）を算出し、この値の大きいものから選択する。

【0069】選択されたキーワードは、情報要約として出力部2に表示される。たとえば、キーワード「米朝高官協議」、「北朝鮮」、「核問題」が選択された場合、

図6のような出力結果が得られる。

【0070】なお、（表4）は、単位1から取り出されたキーワードのみでなく、全ての単位から取り出されたキーワードからなっているが、情報要約として取り出す際には一つの単位から切り出されたキーワードの中だけで単位間共有度を比較し、その中で最も単位間共有度の大きいものから順にキーワードを取り出す方法が考えられる。これにより、一つの単位の中で、他の単位でも共通に扱われているキーワードを取り出すことができる。

【0071】次に、第1の発明の実施の一形態を他の例題に当てはめた場合の処理について説明する。ここで説明する例題は、入力部1から入力されるデータが英語で記述されたものである。例えば、図37のデータ例を考える。このデータ例の各単位は、Los Angeles Times, Chicago Tribune, USA TODAY, Atlanta Constitutionなどの記事に対する概要（Abstract）をデータとして取り

出したものである。このような英語のデータ例に対して処理を行う場合について日本語と異なる部分について特に詳しく説明する。フローチャートは日本語の場合と同じく図3のようになる。

【0072】ステップa1～ステップa2では、日本語に対する処理と同様である。

【0073】ステップa3では、日本語では字種の違いに基づいてキーワードを切り出したのに対し、英語の場合には空白ですでに単語が区切られているので、空白を切れ目として切り出す。したがって、日本語に比べて簡単な処理で行える。ただし、本実施の一形態では大文字で始まっているものは1文字以上、小文字で始まっているものは5文字以上の単語のみをキーワードとして切り出すようにする。これにより、冠詞や前置詞をキーワードとして切り出さないようにするためである。また、大文字で始まる単語が連続する場合には、それらをつなげ*

*て一つの単語と見なす。たとえば、“Rosa Mota of Portugal”からは、“Rosa Mota”と“Portugal”がキーワードとして切り出される。このように、大文字で始まる単語の連続を一固まりと見なすのは、人名などを固まりとして扱うためである。例として、図37の単位1の第1段落からは、“U.S. Olympic”, “champion”, “Joan Benoit Samuelson”, “featured”, “NYC Marathon”)が切り出される。ただし、図37の入力データ例では各単位は空白行で段落が分かれているものとする。

10 【0074】ステップa4では、日本語の場合と同様の処理を行う。キーワードカウント結果としては、たとえば図37の単位1から切り出したキーワードから(表5)のような表を得る。

【0075】

【表5】

キーワード	出現頻度
Olympic	2
champion	1
Joan Benoit Samuelson	1
featured	1
marathon	2
Rosa Mota	1
Portugal	1

【0076】(表5)では、“Olympic”の出現頻度が2になっている。これは、“U.S. Olympic”と“Olympic”が別々に切り出されているが、日本語の場合と同様に、“Olympic”が“U.S. Olympic”の一部になっており、かつその部分が5割よりも長いために一つにまとめたからである。“marathon”についても“NYC Marathon”と“marathon”が1つにまとめられた結果、出現頻度が2になっている。また、このようにキーワードが別のキーワードの部

分になっているか否かを判定したり、キーワードどうしが等しいか否かを判定する際には大文字と小文字を区別しない。

【0077】ステップa5～ステップa9では、日本語の場合と同じ処理を行う。たとえば、キーワード頻度表として以下の(表6)のような表が得られる。

40 【0078】

【表6】

キーワード	単位内頻度	単位間共有度
Olympic	6	3
champion	1	0
Joan Benoit Samuelson	1	0
featured	1	0
marathon	7	3
Rosa Mota	3	2
Portugal	3	2

【0079】ステップa10では、日本語の場合と同じ処理を行う。例えば（表6）のようなキーワード頻度表から単位間共有度の大きいキーワードから順に所定数のキーワードを取り出す。たとえば、（表6）から3個取り出す場合には、“Olympic”、“marathon”、“Rosa Mota”が取り出される。ただし、“Portugal”と“Rosa Mota”は単位間共有度も単位内頻度も等しいために、空白も含めた文字数として“Rosa Mota”が9文字、“Portugal”が8文字であり、文字数の長い方を選択している。選択したキーワードは図38に示すように出力部2に出力する。日本語では図6に示すように「情報要約」をタイトルにしているが、英語では“Trend”をタイトルにしている。

【0080】次に、第2の発明にかかる実施の一形態を図を参照しながら説明する。

【0081】第2の発明の実施の一形態として情報要約方法を用いた文字放送受信装置を示す。図7は、第2の発明の実施の一形態のシステム構成図である。図7において、21は文字放送を受信する文字放送受信部、22は所定の番組のチャンネルを記憶するチャンネル記憶部、23はチャンネル記憶部22に記憶されているチャンネルの各番組を文字放送受信部21で受信し、受信した各番組ごとに番組に含まれているキーワードを抽出する文字放送キーワード切り出し部、24は多くの番組にまたがって現れているキーワードほど点数が高くなるようにキーワードごとに点数を算出する文字放送点数算出部、25は算出した点数に基づいてキーワードを選択し、情報要約として抽出する文字放送要約部、26は要約結果を表示する表示部である。

【0082】以上のように構成されたシステムが実行されるハードウェア構成を図8に示す。図8は、基本的に汎用の計算機システムの構成と同じであり、図2で示した第1の発明の実施の一形態のハードウェア構成の構成部分と図7で示した第2の発明の実施の一形態のシステ

ム構成の構成部分からなっているために、同一構成部分に対して同一番号を付し、詳細な説明は省略する。

【0083】以上のように構成された文字放送受信装置の動作を図9のフローチャートを参照しながら説明する。また、本発明は文字放送受信方式自体に関わるものではないので文字放送受信部分については文字放送受信部21といった一構成部分としてまとめ、詳細な説明は省略する。

【0084】ステップb1～ステップb2では、チャンネル記憶部22に記憶されているチャンネルから一つを取り出し、取り出したチャンネルに従って文字放送受信部21で放送を受信する。ステップb1では、ステップが呼び出される度にチャンネル記憶部22に記憶されているチャンネルを順に取り出すために、記憶されているチャンネルがN個であればN+1回呼び出された時点で取り出すべきチャンネルがなくなる。この場合、フローチャートに示すようにステップb5へ進む。チャンネル記憶部22に記憶されているチャンネルは以下の（表7）に示すようになっている。

【0085】

【表7】

チャンネル	
4 ch	02#
6 ch	101#
:	:

【0086】チャンネル記憶部22に（表7）のようなデータが記憶されている場合、初めにステップb1が呼び出された時には4chと02#が取り出される。このようなチャンネルになっているのは、現在の文字放送においてチ

チャンネルの数字と#が付く数字で番組を指定するようになっているためである。また、ステップb2では受信した文字放送の番組の内容を主記憶装置12に記憶する。ただし、記憶する場所は外部記憶装置13であっても構わない。チャンネル記憶部22に記憶させるチャンネルの例としては、ニュース番組のチャンネルを複数記憶させる、スポーツ番組のチャンネルを複数記憶させるなど、特定のジャンルの番組が放送されているチャンネルを記憶させることにより、本実施の一形態が特定のジャンルの最新のトレンドを要約結果として取り出すことができる。

【0087】ステップb3では、文字放送キーワード切り出し部23においてステップb2で記憶された文字放送の番組の内容からキーワードを切り出す。キーワードの切り出しは、第1の発明の実施の一形態のステップa3と同様の方法で行う。文字放送で受信した番組1個分が第1*

キーワード	番組内頻度	番組間共有度

【0091】ただし、処理内容は第1の発明の実施の一形態と同様であり、単位内頻度に対して行った処理と同じ処理を番組内頻度に対して行い、単位間共有度に対して行った処理と同じ処理を番組間共有度に対して行う。また、第1の発明の実施の一形態では点数算出部4で実行していた部分は、本実施の一形態では文字放送点数算出部24で実行する。

【0092】ステップb10では、文字放送要約部25において第1の発明の実施の一形態のステップa10と同様の処理を行う。第1の発明の実施の一形態の説明で用いた図5のデータ例は文字放送から取り出したものであり、出力結果の例としては第1の発明の実施の一形態で示した図6と同様である。出力結果は表示部26に表示される。

【0093】なお、第2の発明の実施の一形態を英語の例題に当てはめた場合の処理については、第1の発明の実施の一形態において説明した内容と同様であるためにここでは省略する。本発明においても英語で記述されたデータを扱えることは第1の発明と同様である。

【0094】次に、第3の発明にかかる実施の一形態を図面を参照しながら説明する。

【0095】第3の発明にかかる実施の一形態としてキーワード間の関連を考慮した情報要約方法、および情報要約装置を示す。図10は、第3の発明の実施の一形態のシステム構成図である。図10には第1の発明の実施の一形態の一構成例と同一の構成部分があるので、同一

*の発明の実施の一形態の1単位に対応する。

【0088】ステップb4では、ステップb3で切り出したキーワードに対して、第1の発明の実施の一形態のステップa4と同様の方法でキーワードの数を数えてスタックに記憶（プッシュ）する。キーワードの数を数えた結果は第1の発明の実施の一形態で示したカウント結果表である（表1）と同様の形式にしてスタックに記憶する。

【0089】ステップb5～ステップb9では、第1の発明の実施の一形態のステップa5～ステップa9と同様の処理を行う。ただし、第1の発明の実施の一形態では、キーワード頻度表の形式が（表2）のようになっているが、本実施の一形態では単位の変わりに文字放送の番組を用いているために、項目名を以下の（表8）のようにする。

【0090】

【表8】

構成部分に対しては同一番号を付して詳細な説明は省略する。図10において、31はキーワード切り出し部3で抽出したキーワードに対して同一の単位中の同一段落から得られたキーワードどうしを関連付けるキーワード関連付け部、32はキーワード切り出し部3で切り出されたキーワード、およびキーワード関連付け部31で関連づけられたキーワードどうしの関連付けに対して、より多くの単位にまたがって現れているキーワード、およびキーワードどうしの関連付けほど点数が高くなるように点数を算出する関連付点数算出部である。33は関連付点数算出部32で算出された点数に基づいてキーワード、およびキーワードどうしの関連付けを選択し、情報要約として出力部2に表示する関連付要約部である。

【0096】以上のように構成されたシステムが実行されるハードウェア構成は図2に示した第1の発明の実施の一形態のハードウェア構成と同じである。したがって、詳細な説明は省略する。

【0097】以上のように構成されたキーワード間の関連を考慮した情報要約方法、および情報要約装置の動作を図11のフローチャートを参照しながら説明する。

【0098】ステップc1～ステップc3は、第1の発明の実施の一形態のステップa1～ステップa3と同様の処理を行う。

【0099】ステップc4では、キーワード関連付部31において同一単位中の同一段落のキーワードどうしを関連づける。キーワードどうしの関連付けは、2つのキー

ワードの組を抽出する。キーワードどうしの関連付けを表すキーワードの組においては、順序関係は無いものとする。すなわち、(Keyword1, Keyword2)と(Keyword2, Keyword1)を同一の関連付けと見なす。ただし、別々に切り出された2つのキーワードのうち片方が、もう片方の部分になっている場合には、ステップa4の説明で述べたのと同様の方法で同一とみなし、一つにまとめてから関連付けを行う。たとえば、図5に示したデータ例の単位1では、空白行ごとに段落が区切られている。このデータ例の単位1の第1段落から抽出されるキーワード「泉佐野」、「ヘリ墜落」、「大阪府警」、「朝日新聞社取材用」、「ヘリコプター」に対して、キーワードどうしの関連付け(泉佐野, ヘリ墜落)、(泉佐野, 大阪府警)、(泉佐野, 朝日新聞社取材用)、(泉佐野, ヘリコプター)、(ヘリ墜落, 大阪府警)、(ヘリ墜落, 朝日新聞社取材用)、(ヘリ墜落, ヘリコプター)、(大阪府警, 朝日新聞社取材用)、(大阪府警, ヘリコプター)、*

* (朝日新聞社取材用, ヘリコプター) が生成される。ただし、データ例からは「泉佐野」と「泉佐野市」がキーワードとして抽出されるが、前述のとおりステップa4の説明で述べた方法により一つのキーワード「泉佐野」にまとめている。

【0100】ステップc5では、ステップc3、ステップc4で抽出したキーワードとキーワードどうしの関連付けをカウント表の形式にし、第1の発明の実施の一形態のステップa4と同様にFILO(First In Last Out)のスタックに記憶(プッシュ)する。その際には、同一のキーワード、キーワードどうしの関連付けが複数個ある場合には、それらをまとめて、数を出現頻度の項目に記憶する。カウント表の形式は、以下の(表9)の形式である。

【0101】

【表9】

キーワード、あるいは関連付け	出現頻度

【0102】(表9)の項目「キーワード、あるいは関連付け」には、具体的なキーワード、あるいはキーワードどうしの関連付けが項目値として入る。キーワードどうしの関連付けの場合は、2つのキーワードの組になるために(Keyword1, Keyword2)の形式の項目値になる。たとえば、図5に示したデータ例の単位1の第1段落に

において、抽出されるキーワード、キーワード間の関連付けから得られるカウント表は、以下の(表10)のようになる。

【0103】

【表10】

キーワード、あるいは関連付け	出現頻度
泉佐野	2
ヘリ墜落	1
大阪府警	1
朝日新聞社取材用	1
ヘリコプター	1
(泉佐野,ヘリ墜落)	2
(泉佐野,大阪府警)	2
(泉佐野,朝日新聞社取材用)	2
(泉佐野,ヘリコプター)	2
(ヘリ墜落,大阪府警)	1
(ヘリ墜落,朝日新聞社取材用)	1
(ヘリ墜落,ヘリコプター)	1
(大阪府警,朝日新聞社取材用)	1
(大阪府警,ヘリコプター)	1
(朝日新聞社取材用,ヘリコプター)	1

【0104】(表10)のキーワードどうしの関連付けの出現頻度は、関連付けされている2つのキーワードの出現頻度のうち大きい方の値にする。

【0105】ステップc6では、ステップc5でスタックに記憶したキーワード、あるいはキーワードどうしの関連付けと出現頻度の組で表されるデータを1つ取り出す。

【0106】ステップc7では、ステップc6で取り出したデータの項目「キーワード、あるいは関連付け」の値を取り出し、その値がキーワードである場合にはステップc8へ進み、キーワードどうしの関連付けの場合にはステップc11へ進む。取り出した項目値がキーワードかキーワードどうしの関連付けかを判断する手法は、項目値が*40

* (keyword1,keyword2) のように2つのキーワードの組になっている場合には、キーワードどうしの関連付けと判断し、Keyword1のように単純にキーワードが1つだけある場合には、キーワードと判断する。

【0107】ステップc8～ステップc10では、関連付点数算出部32において第1の発明の実施の一形態のステップa7～ステップa9と同様の処理を行う。ただし、キーワード頻度表の形式は、以下に示す(表11)のようになっている。

【0108】

【表11】

キーワード、あるいは関連付け	単位内頻度	単位間共有度

【0109】(表11)のキーワード頻度表は、(表 50 2)に示した表と類似の形式であるが、項目名「キーワ

ード」が項目名「キーワード、あるいは関連付け」に変わっている。これは、この項目値としてキーワードのみでなく、キーワードどうしの関連付けも格納されるためである。ただし、ステップc8～ステップc10では、ステップc6で取り出したデータの項目「キーワード、あるいは関連付け」の値がキーワードである場合に対してのみ処理するために、項目名が異なるだけでステップa7～ステップa9と同様に処理を進める。

【0110】ステップc11が実行されるのは、ステップc6で取り出されたデータがキーワードどうしの関連付けに対するデータである場合であり、その関連付けと同一の関連付けがキーワード頻度表に存在するか否かを判定する。キーワード表にステップc6で取り出したキーワードどうしの関連付けと同一のものがあるか否かを判定する方法は、キーワード頻度表の項目「キーワード、あるいは関連付け」の値と順に比較して実行する。ただし、判定の方法は効率のよい別の方法であっても構わない。また、キーワードどうしの関連付けの場合の判定時には、ステップc4の説明で述べた通り関連付けの表現形式(Keyword1, Keyword2)に対して、(Keyword2, Keyword1)であっても同一と見なす。また、各キーワードについて、片方のキーワードがもう片方のキーワードの部分になっている場合には、ステップa7で説明したのと同様の基準と方法で一つにまとめる。

【0111】ステップc12は、ステップc6で取り出したキーワードどうしの関連付けがキーワード頻度表の中に既に登録されている場合に実行される。これは、複数の異なる段落中に同一のキーワードどうしの関連付けが存在した場合である。この場合には、キーワード頻度表の単位間共有度の値に1を加える。実際にはキーワードどうしの関連付けの場合、単位ではなく段落で扱うために、項目「単位間共有度」、「単位内頻度」は、項目名を「段落間共有度」、「段落内頻度」とするべきであるが、キーワードの場合と同一のキーワード頻度表を用い

るために、項目名はキーワードの場合に合わせている。また、単位内頻度の値はステップc6で取り出した組の出現頻度を単位内頻度の項目の値に加える。たとえば、ステップc6で取り出された組が（（米朝高官協議，北朝鮮），2）である場合を考える。この場合、キーワード頻度表の中の属性「キーワード、あるいは関連付け」の項目値が（米朝高官協議，北朝鮮）であるデータの単位内頻度の値を2増やし、単位間共有度の値を1増やす。

【0112】ステップc13は、キーワード頻度表の中にステップc6で取り出したキーワードどうしの関連付けと同一のものが存在しない場合に実行される処理であり、この場合にはステップc6で取り出した組のキーワードどうしの関連付けをキーワード頻度表に追加する。追加する際には、単位内頻度はステップc6で取り出した組の出現頻度に設定する。また、単位間共有度は0にする。以上のステップc11～ステップc13は関連付点数算出部32において実行される。

【0113】ステップc14では、関連付要約部33において、キーワード頻度表のキーワードの中から単位間共有度が大きいものから順に所定数のキーワードを取り出す。次に、それぞれのキーワードに対して、そのキーワードを含む関連付けのうち単位間共有度の大きいものから順に所定数の関連付けを取り出す。最後に、検索した結果を出力部2に表示する。ただし、単位間共有度が同じ場合には単位内頻度の大きいものを優先する。単位内頻度まで同じ場合には、文字数の多いキーワード、あるいは文字数の多いキーワードを含むキーワードどうしの関連付けを優先して取り出す。たとえば、（表12）のようなキーワード頻度表に対して、キーワードを2個と、各キーワードに対して関連付けを2個ずつ取り出す場合を考える。

【0114】

【表12】

キーワード、あるいは関連付け	単位内頻度	単位間共有度
泉佐野	3	1
朝日新聞社	5	3
ヘリ墜落	3	3
毎日新聞社	1	0
(泉佐野, 朝日新聞社)	4	1
(泉佐野, ヘリ墜落)	2	0
(泉佐野, 毎日新聞社)	1	0
(朝日新聞社, ヘリ墜落)	3	2
(朝日新聞社, 毎日新聞社)	2	0
米朝高官協議	6	4
北朝鮮	4	3
調印式	1	0
核問題	3	2
燃料棒	1	0
(米朝高官協議, 北朝鮮)	4	3
(米朝高官協議, 調印式)	1	0
(米朝高官協議, 核問題)	3	2
(米朝高官協議, 燃料棒)	1	0
(北朝鮮, 調印式)	1	0
(北朝鮮, 核問題)	2	1
(核問題, 燃料棒)	1	0
平和条約	2	1
仮調印	2	1
イスラエル	1	0
ヨルダン	1	0
(平和条約, 仮調印)	2	1
(平和条約, イスラエル)	1	0
(平和条約, ヨルダン)	1	0
(仮調印, イスラエル)	1	0
(仮調印, ヨルダン)	1	0
(イスラエル, ヨルダン)	1	0

【0115】この場合、単位間共有度の大きいものから2個のキーワードを取り出すと「米朝高官協議」、「朝日新聞社」になる。キーワード「朝日新聞社」とキーワード「ヘリ墜落」とキーワード「北朝鮮」は単位間共有度が等しいが、単位内頻度に基づいて「朝日新聞社」が取り出される。また、各キーワードを含むキーワードどうしの関連付けを単位間共有度に基づいて取り出すと、(米朝高官協議, 北朝鮮)、(米朝高官協議, 核問題)、(朝日新聞社, ヘリ墜落)、(泉佐野, 朝日新聞社)にな

50

り、関連付けられているキーワードは「北朝鮮」、「核問題」、「ヘリ墜落」、「泉佐野」になる。このようにして取り出したキーワード、およびキーワードどうしの関連付けの表示例を図12に示す。図12においては3つのキーワードの組合せで1つの話題を表している。なお、キーワード、およびキーワードどうしの関連付けを選択する際には、別の方法として単位内頻度と単位間共有度の重み付け和の大きいもの順に所定数を検索する方法も考えられる。重み付け和は、予め重み付けを表す定

数 S, T を定め、各キーワードごとに（単位内頻度 $\times S$ ＋単位間共有度 $\times T$ ）を算出する。

【0116】また、ステップc14において要約結果を取り出す別の方法として、より話題を正確に取り出す方法を以下に述べる。まず、単位間共有度の最も大きいキーワードを選択してAとし、キーワードAと関連付けられているキーワードのうち、Aとのキーワードどうしの関連付けの単位間共有度が最も大きいキーワードを選択してキーワードBとする。次に、選択されたキーワードA、キーワードBの両方に関連付けられていて、それぞれのキーワードとのキーワードどうしの関連付けの単位間共有度を合計した値が最も大きいキーワードを選択してCとする。そして、キーワードA、B、Cを組み合わせる1つ目の話題とする。2つ目の話題を取り出す際には、1つ目の話題を表すキーワードと、そのキーワードを含むキーワードどうしの関連付けをキーワード頻度表から消去し、その後と同じ処理を行う。また、選択の際に単位間共有度が等しい複数のキーワード、あるいはキーワードどうしの関連付けがある場合は、上述のように単位間頻度やキーワードの長さなどを比べる。

【0117】例として（表12）から2つの話題を取り出す際の処理について述べる。1つ目の話題を取り出す際には、キーワードAとして「米朝高官協議」が選択され、このキーワードとのキーワードどうしの関連付けの単位間共有度が最も大きい「北朝鮮」がキーワードBとして選択される。さらに、これらのキーワードの両方と関連付けられており、かつキーワードどうしの関連付けの単位間共有度を合計した値が最も大きい「核問題」が

キーワードCとして選択される。2つ目の話題を取り出す際には、1つ目の話題に用いたキーワード「米朝高官協議」、「北朝鮮」、「核問題」と、これらのキーワードを含むキーワードどうしの関連付けをキーワード頻度表から消去した後に同様の方法でキーワードA、B、Cを選択する。（表12）からは、2つ目の話題を表すキーワードとして「朝日新聞社」、「ヘリ墜落」、「泉佐野」が選択される。

【0118】次に、第3の発明の実施の一形態を他の例題に当てはめた場合の処理について説明する。ここで説明する例題は、入力部1から入力されるデータが英語で記述されたものである。例えば、第1の発明の実施の一形態において説明したのと同様に図37のデータ例を考える。英語のデータ例に対して処理を行う場合について日本語と異なる部分について特に詳しく説明する。フローチャートは、日本語の場合と同じく図11のようになる。

【0119】ステップc1～ステップc2では、日本語に対する処理と同様である。

20 【0120】ステップc3での日本語に対する処理との相違点については、第1の発明の実施の一形態のステップa3を英語の例題に当てはめた際に説明したのと同じである。

【0121】ステップc4～ステップc14では、日本語に対する処理と同様である。ステップc5で得られるカウンタ表の例を（表13）に示す。

【0122】

【表13】

キーワード、あるいは関連付け	出現頻度
Olympic	2
champion	1
Joan Benoit Samuelson	1
featured	1
marathon	2
Rosa Mota	1
Portugal	1
(Olympic, champion)	2
(Olympic, Joan Benoit Samuelson)	2
(Olympic, featured)	2
(Olympic, marathon)	2
(Olympic, Rosa Mota)	2
(Olympic, Portugal)	2
(champion, Joan Benoit Samuelson)	1
(champion, featured)	1
(champion, marathon)	2
(Joan Benoit Samuelson, featured)	1
(Joan Benoit Samuelson, marathon)	2
(featured, marathon)	2
(marathon, Rosa Mota)	2
(marathon, Portugal)	2
(Rosa Mota, Portugal)	1

【0123】また、キーワード頻度表の例を（表14）に示す。このようなキーワード頻度表から、ステップc1 4において、まず単位間共有度が最も大きいキーワードを選択してAとし、キーワードAと関連付けられているキーワードのうち、キーワードAとのキーワードどうしの関連付けの単位間共有度が最も大きいものを1つ選択してキーワードBとする。次に、選択されたキーワードA、キーワードBの両方に関連付けられていて、それぞれのキーワードとのキーワードどうしの関連付けの単位間共有度を合計した値が最大のキーワードを選択してCとする。そして、キーワードA、B、Cを組み合わせるとして出力する。ただし、このようなキーワ

ードの組合せで表される話題を複数個取り出しても構わない。たとえば、（表14）で具体的に値が記述されている部分のみからこのような話題を2個取り出すと、{marathon, Rosa Mota, Portugal}、{Olympic, Gelindo Bordin, Italy} が得られる。2個取り出す場合、1個目の話題を表すキーワードと、そのキーワードを含むキーワードどうしの関連付けをキーワード頻度表から消去した後、2個目の話題を同様の処理で取り出す。出力結果としては図39に示すような結果が得られる。

【0124】

【表14】

キーワード、あるいは関連付け	単位内頻度	単位間共有度
Olympic	6	5
champion	1	0
Joan Benoit Samuelson	1	0
:	:	:
(Olympic, champion)	1	0
(Olympic, Joan Benoit Samuelson)	1	0
:	:	:
Rosa Mota	3	2
Portugal	3	2
marathon	7	6
:	:	:
(Rosa Mota, Portugal)	3	2
(Rosa Mota, marathon)	3	2
(Rosa Mota, Olympic)	2	1
(Portugal, marathon)	3	2
:	:	:
Gelindo Bordin	2	1
Italy	2	1
overtakes	1	0
:	:	:
(Gelindo Bordin, Italy)	2	1
(Gelindo Bordin, overtakes)	2	0
(Gelindo Bordin, Olympic)	6	1
(Gelindo Bordin, marathon)	7	1
(Italy, Olympic)	2	1
:	:	:

【0125】なお、本実施の一形態では同一単位中の同一段落で説明したが、段落の代わりに文単位など他の決め方による範囲でも構わない。

【0126】なお、本発明ではキーワードの単位内頻度、単位間共有度の両方を算出し、要約結果として表示するキーワードを選択したが、単位間共有度のみを用いて選択する方法も考えられる。これは、キーワードが同一単位内で繰り返し出現するのは、単なる言い替えであったり表現上の理由である場合があり、単位内頻度はキーワードとしての重要性よりも表現上の理由で大きな

場合があるからである。このため、処理の簡素化を優先して単位内頻度を算出しないで単位間共有度のみで処理を行う方法も必要である。

【0127】また、文書の要約結果としてキーワードを表示する際には、キーワードの組合せによって意味内容が伝わる場合が多く、個々のキーワードの重要性を処理するよりもキーワードどうしの関連付けの重要性のみを扱った方が要約方法としても簡素化され、実際の装置として実現する際に有用になる場合がある。以下、この方法について説明する。

【0128】まず、ステップc5においてキーワードに対する処理を行わずにキーワードどうしの関連付けに対する処理のみを行う。また、出現頻度に関する処理も行わない。すなわち、キーワードについてはスタックに記憶せずに、キーワードどうしの関連付けのみをスタックに記憶する。その際には、出現頻度に関する処理を行わないために、カウント表の形式は(表9)の「出現頻度」の項目を取り除いた形式になる。すなわち、項目「キー*

*ワード、あるいは関連付け」のみからなり、この項目にはキーワードどうしの関連付けのみが項目値として入る表になる。ただし、同一のキーワードどうしの関連付けは一つにまとめる。カウント表の例を(表15)に示す。

【0129】

【表15】

キーワード、あるいは関連付け

(泉佐野,ヘリ墜落)

(泉佐野,大阪府警)

(泉佐野,朝日新聞社取材用)

(泉佐野,ヘリコプター)

(ヘリ墜落,大阪府警)

(ヘリ墜落,朝日新聞社取材用)

(ヘリ墜落,ヘリコプター)

(大阪府警,朝日新聞社取材用)

(大阪府警,ヘリコプター)

(朝日新聞社取材用,ヘリコプター)

【0130】また、ステップc8～ステップc10は不要になり、ステップc7では判定処理を行わずにステップc11へ常に進む。ステップc12、ステップc13では、ステップc6で取り出したデータに項目「出現頻度」が無く、この項目値を用いて行う単位内頻度に対する処理は行わない。このような処理により、たとえば(表16)に示す

30 ように(表12)のキーワード頻度表のうち、キーワードどうしの関連付けのみからなり、かつ項目「単位内頻度」のないキーワード頻度表が得られる。

【0131】

【表16】

キーワード、あるいは関連付け	単位間共有度
(泉佐野, 朝日新聞社)	1
(泉佐野, ヘリ墜落)	0
(泉佐野, 毎日新聞社)	0
(朝日新聞社, ヘリ墜落)	2
(朝日新聞社, 毎日新聞社)	0
(米朝高官協議, 北朝鮮)	3
(米朝高官協議, 調印式)	0
(米朝高官協議, 核問題)	2
(米朝高官協議, 燃料棒)	0
(北朝鮮, 調印式)	0
(北朝鮮, 核問題)	1
(核問題, 燃料棒)	0
(平和条約, 仮調印)	1
(平和条約, イスラエル)	0
(平和条約, ヨルダン)	0
(仮調印, イスラエル)	0
(仮調印, ヨルダン)	0
(イスラエル, ヨルダン)	0

【0132】さらに、ステップc14においてはキーワード頻度表のキーワードどうしの関連付けの中から単位間共有度の最も大きいものを取り出す。たとえば、(表16)からは、(米朝高官協議, 北朝鮮), 3)が取り出される。次に、取り出されたキーワードどうしの関連付けを構成する2つのキーワードA, Bそれぞれについて、Aと関連付けられて、かつBとも関連付けられているキーワードのうちAとの関連付け、Bとの関連付けの単位間共有度の和が最も大きいキーワードを選択する。たとえば、(表16)のキーワード頻度表を考えれば、キーワードA, Bは「米朝高官協議」、「北朝鮮」であり、これらと関連付けられており、かつその関連付けの単位間共有度の和が最大になるのは「核問題」である。したがって、キーワード「核問題」が選択される。キーワードA, Bと、この選択の結果得られたキーワードとを組み合わせて1つの話題として得る。また、さらに別の話題を情報要約として得る際には、キーワード頻度表から既に選ばれたキーワードを含むキーワードどうしの

関連付けを消去し、残ったキーワードどうしの関連付けの中から同様の処理で3つのキーワードを選択する。たとえば、上記の例では、{「米朝高官協議」、「北朝鮮」、「核問題」}が選択された後、これらのキーワードを含むキーワードどうしの関連付けが消去され、次に単位間共有度の大きいキーワードどうしの関連付けとして(朝日新聞社, ヘリ墜落), 2)が取り出される。この場合、キーワードA, Bは、「朝日新聞社」、「ヘリ墜落」になり、これらと関連付けられていて、そのキーワードどうしの関連付けの単位間共有度の和が最大になるのは「泉佐野」であり、このキーワードが選択される。この結果、(表16)からは、情報要約として{「米朝高官協議」、「北朝鮮」、「核問題」}と{「朝日新聞社」、「ヘリ墜落」、「泉佐野」}が取り出され、図12のような出力結果が得られる。

【0133】また、このように単位間共有度のみを用いる方法を英語の例題に適用した場合については、上述のような日本語処理との相違点があるのみで、適用できる

ことは言うまでもない。

【0134】次に、第4の発明にかかる実施の一形態を図面を参照しながら説明する。

【0135】第4の発明の実施の一形態としてキーワード間の関連を考慮した情報要約付き文字放送受信装置を示す。図13は、第4の発明の実施の一形態のシステム構成図である。図13には第2の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図13において、41は文字放送キーワード切り出し部23で抽出したキーワードに対して同一番組中の同一段落から得られたキーワードどうしを関連付ける文字放送キーワード関連付け部、42はより多くの番組にまたがって現れているキーワード、およびキーワードどうしの関連付けほど点数が高くなるようにキーワード、およびキーワードどうしの関連付けごとに点数を算出する文字放送関連付け点数算出部、43は文字放送関連付け点数算出部で算出した点数に基づいてキーワード、およびキーワードどうしの関連付けを選択し、表示部26に表示する文字放送関連付け要約部である。

【0136】以上のように構成されたシステムが実行されるハードウェア構成は図8に示した第2の発明の実施の一形態のハードウェア構成と同じである。したがって、詳細な説明は省略する。

【0137】以上のように構成されたキーワード間の関連を考慮した情報要約機能付き文字放送受信装置の動作を図14のフローチャートを参照しながら説明する。

【0138】ステップd1～ステップd3では、第2の発明の実施の一形態のステップb1～ステップb3と同様の処理を行う。

【0139】ステップd4～ステップd14では、第3の発明の実施の一形態のステップc4～ステップc14と同様の処理を行う。ただし、ステップd4は文字放送キーワード関連付け部41において実行される。ステップd8～ステップd13は文字放送関連付け点数算出部42において実行される。ステップd14は、文字放送関連付け要約部43において実行される。

【0140】なお、本発明においても第3の発明と同様に、単位間共有度のみを用いて選択する方法が考えられる。この方法については、ステップd4～ステップd14がステップc4～ステップc14と同様の処理を行うことから第3の発明で説明したのと同様の変更で実現できるので詳細な説明は省略する。また、第3の発明の実施の一形態で説明したのと同様にキーワードどうしの関連付けのみに対して処理を行う方法が考えられる。

【0141】次に、第5の発明にかかる実施の一形態を図面を参照しながら説明する。

【0142】第5の発明の実施の一形態として情報要約方法、および情報要約装置を示す。図16は、第5の発明の実施の一形態のシステム構成図である。図16には

第1の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図16において、第1の発明の実施の一形態のシステム構成例と異なるのは、キーワード切り出し部3で切り出したキーワード間の類似性を算出する類似度付点数算出部51である。

【0143】以上のように構成されたシステムが実行されるハードウェア構成は図2に示した第1の発明の実施の一形態のハードウェア構成と同じである。

【0144】以上のように構成された情報要約方法、および情報要約装置の動作を図17のフローチャートを参照しながら説明する。

【0145】ステップe1～ステップe3では、第1の発明の実施の一形態のステップa1～ステップa3と同様の処理を行う。

【0146】ステップe4～ステップe9は類似度付点数算出部51で実行される。

【0147】ステップe4では、ステップe3で切り出したキーワードに対して、類似度を考慮しながら出現頻度を算出する。出現頻度の算出の際には、まず完全に一致するキーワードをまとめて頻度を抽出された個数にする。たとえば、「国際平和維持軍」が4個切り出された場合、キーワード「国際平和維持軍」を1つにまとめて出現頻度を4にする。次に、同一ではない2つのキーワードA、Bに対しては、文字と順序が一致している文字数をキーワードA、Bの文字数の平均で割った値を類似度として用いる。たとえば、「米朝高官会議」と「米朝高官協議」であれば「米朝高官議」が一致しているために文字と順序が一致している文字数は5文字になる。2つのキーワードの文字数の平均は6文字である。この結果、2つのキーワードの類似度は、 $5/6=0.83$

(小数点以下2桁まで)になる。さらに、2つのキーワードA、Bの間で算出した類似度をSとして、Aの出現頻度(S×Bの出現頻度)を加え、Bの出現頻度(S×Aの出現頻度)を加える。たとえば、類似度による出現頻度の加算を行う前の「米朝高官会議」の出現頻度が3で、「米朝高官協議」の出現頻度が5であった場合、類似度0.83を用いて「米朝高官会議」の出現頻度は $3+5\times 0.83=7.15$ になり、「米朝高官協議」の出現頻度は $5+3\times 0.83=7.49$ になる。このようなキーワード間の類似度は、ステップe3で切り出されたキーワードすべての組み合わせに対して算出する。ただし、計算時間を少なくするために2つのキーワードで一致する文字数が5割以上のもののみに対して類似度を算出するなどの方法も考えられる。このように、類似度に基づく出現頻度を算出するために、たとえば以下の(表17)のように整数以外の出現頻度を含むキーワードカウント結果を得る。

【0148】

【表17】

キーワード	出現頻度
泉佐野	1. 8 6
ヘリ墜落	1
大阪府警	1
泉佐野市	1. 8 6
朝日新聞社取材用	1
ヘリコプター	1
新聞大会開	1. 6 2
第47回新聞大会	1. 6 2
京都市	1
新聞販売	1
正常化	1
特別宣言	1
電動茶	2
京都府茶協同組合	1
米朝高官協議合意	1. 7 1
北朝鮮	1
核問題	1
米朝高官会議	1. 7 1
ジュネーブ	1
調印式	1

【0149】また、得られたキーワードカウント結果はFILO(First In Last Out)のスタックに単位を識別する識別子と組み合わせて記憶(プッシュ)する。第1の発明の実施の一形態では、ステップa4において同一と見なされるキーワードは1つにまとめられ、かつキーワードの類似度を考えない。このため、後のステップa5〜ステップa9でキーワード頻度表の単位内頻度、単位間共有度を更新する際に、どの単位から抽出されたキーワードかをスタックに記憶していなくても、同じ単位から抽出されたキーワードどうしは同一と見なされることはない。この結果、同じ単位から抽出されたキーワードどうしが、ステップa7で同一と見なされてステップa8で単位間共有度の更新処理を受けるといったことは生じない。これに対し、本実施の一形態のように類似度を考慮しているにもかかわらず、さらにスタックに抽出された単位

の識別子を記憶しなければ、キーワード頻度表を更新する際に同じ単位から抽出されたキーワードどうしも、他の単位から抽出されたキーワードどうしも区別なく類似度が算出される。この結果、別の単位から抽出されたキーワードか同じ単位から抽出されたキーワードかの区別ができなくなり、後述のステップe7における頻度修正処理で別の単位に同様のキーワードが出現していることを表す単位間共有度を算出することができなくなる。したがって、第1の発明の実施の一形態と異なり、スタックには単位を識別する識別子を組み合わせて記憶する。なお、この識別子は他の単位と区別できればよいので、処理の中でb1, b2, ... などのように生成して割り当てればよい。

【0150】ステップe5では、第1の発明の実施の一形態のステップa5と同様の処理を行う。ただし、ステップ

e5で取り出されるデータは、ステップa5と比較して、単位識別子が組み合わされている点が異なる。

【0151】ステップe6では、ステップe5で取り出したデータの中の単位を識別する識別子と変数Bidの内容を比較する。ただし、変数Bidの初期値としては、いずれの識別子とも等しくなると同時に、比較対象の値を持つように設定する。たとえば、単位の識別子としては0より大きい整数を割り当て、Bidの初期値を-1とし、このステップでの比較処理において、Bidが-1の時のみ、Bidに比較対象の識別子をまず代入してから比較処理を行う。比較の結果、等しい場合には、ステップe7へ進み、それ以外の場合にはステップe11へ進む。

【0152】ステップe7では、第1の発明の実施の一形態のステップa6と同様の処理を行う。

【0153】ステップe8では、第1の発明の実施の一形態で示した(表2)のようなキーワード頻度表のキーワードの項目に対して、まずステップe7で取り出したキーワードと類似度を算出する。また、類似度の算出の際には、後述のステップe9で同一のキーワードがキーワード頻度表の中に存在するか否かを判定するために、フラグ*20

*変数などを用いて同一のキーワード(類似度=1)が存在した場合には、ステップe9で判定可能なようにする。次に、類似度が0.5以上の場合に、キーワード頻度表の単位内頻度と単位間共有度を更新する。単位内頻度、単位間共有度の更新方法は、以下の方法で行う。ステップe5で取り出したデータのキーワードの項目値をA、出現頻度の項目の値をFとし、キーワード頻度表のキーワードの項目値をB、単位内頻度をFi、単位間共有度をFeとする。キーワードA、Bの類似度をSとした場合に、キーワード頻度表の単位内頻度の値をFi+S×Fに、単位間共有度の値をFe+Sに更新する。たとえば、ステップe5で取り出したデータが、(キーワード、出現頻度)=(米朝高官会談, 2.63)であり、以下の(表18)のようなキーワード頻度表に対して更新処理をする場合を考える。ただし、この例で用いている数値は説明のために定めた仮のものであり実際のデータから算出したものではない。

【0154】

【表18】

キーワード	単位内頻度	単位間共有度
電動茶	2. 3 2	1. 8 7
京都府茶協同組合	1. 6 7	0
米朝高官協議	4. 1 5	2. 7 1
米朝高官会議	3. 3 4	2. 6 8
:	:	:

【0155】(表18)の中で、キーワード「米朝高官会談」と類似度0.5以上のキーワードは、「米朝高官協議」と「米朝高官会議」である。「米朝高官会談」と「米朝高官協議」の類似度は0.67、「米朝高官会談」と「米朝高官会議」の類似度は0.83である。

(小数点以下第3位を四捨五入)この結果、(表18)のキーワード頻度表の(キーワード、単位内頻度、単位間共有度)={ (米朝高官協議, 4.15, 2.71), (米朝高官会議, 3.34, 2.68) } が { (米朝高官協議, 4.15+0.67×2.63, 2.71+0.67), (米朝高官会議, 3.34+0.83×2.63, 2.68+0.83) } に更新される。

【0156】なお、本実施の一形態ではキーワード頻度表の中のキーワードのうち、ステップe5で取り出したデータと0.5以上の類似度になるものに対してのみ更新処理を施すことにしたが、他の基準でも構わない。類似

度0.5以上のものに対してのみ処理を施したのは、更新処理の回数を少なくするためである。また、図17に示すフローチャートを実行開始してから初めてステップe8を実行する際には、キーワード頻度表は空であり、この場合ステップe8ではなにも行わずにステップe9へ進む。

【0157】ステップe9では、ステップe8でキーワード頻度表の更新処理をする中で、ステップe7で取り出したキーワードと同一のキーワードがキーワード頻度表の中に存在したか否かを判定し、存在した場合にはステップe5へ戻り、存在しなかった場合にはステップe10へ進む。

【0158】ステップe10では、ステップe5で取り出したデータをスタックなどの記憶領域に一時的に記憶する。ただし、ステップe4で使用したものととは別のスタックに記憶する。

【0159】ステップe11では、ステップe10でスタック

などの記憶領域に記憶したデータの一つずつ取り出してキーワード頻度表に追加する。各データをキーワード頻度表に追加する方法は、第1の発明の実施の一形態のステップa9と同様の処理で行う。また、このステップではデータを取り出して記憶領域から消去する。このため、このステップを実行し終わる度にステップe10でスタックに記憶したデータは一度空になる。

【0160】ステップe12では、変数Bidにステップe5で読み出したデータの単位を識別する識別子を代入する。

【0161】ステップe13では、ステップe11と同じ処理を行う。

【0162】ステップe14では、キーワードの出現頻度、単位内頻度、単位間共有度の値が整数ではなく、小数点以下のある実数になっている点が異なるが、第1の発明の実施の一形態のステップa10と同様の処理を行う。

【0163】次に、第6の発明にかかる実施の一形態を図面を参照しながら説明する。

【0164】第6の発明の実施の一形態として情報要約付き文字放送受信装置を示す。図18は、第6の発明の実施の一形態のシステム構成図である。図18には第2の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図18において、61は文字放送キーワード切り出し部23で切り出したキーワード間の類似性を算出する文字放送類似度付点数算出部である。

【0165】以上のように構成されたシステムが実行されるハードウェア構成は図8に示した第2の発明の実施の一形態のハードウェア構成と同じである。したがって、詳細な説明は省略する。

【0166】以上のように構成されたキーワード間の類似性を考慮した情報要約機能付き文字放送受信装置の動作を図19のフローチャートを参照しながら説明する。

【0167】ステップf1～ステップf3では、第2の発明の実施の一形態のステップb1～ステップb3と同様の処理を行う。

【0168】ステップf4～ステップf13は、文字放送類似度付点数算出部61で実行する。

【0169】ステップf4～ステップf13は、第5の発明の実施の一形態のステップe4～ステップe13と同様の処理を行う。ただし、キーワード頻度表の項目名は第2の発明の実施の一形態と同様であり、第5の発明と比べて単位内頻度に対して行った処理と同じ処理を番組内頻度に対して行い、単位間共有度に対して行った処理と同じ処理を番組間共有度に対して行う。

【0170】ステップf14では、キーワードの出現頻度、単位内頻度、単位間共有度の値が整数ではなく、小数点以下のある実数になっている点が異なるが、第2の発明の実施の一形態のステップb10と同様の処理を行う。

【0171】次に、第7の発明にかかる実施の一形態を図面を参照しながら説明する。

【0172】第7の発明の実施の一形態としてキーワードの類似度とキーワード間の関連付けの両方を考慮した情報要約方法、および情報要約装置を示す。図20は、第7の発明の実施の一形態のシステム構成図である。図20には第3の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図20において、71はキーワード間の類似性をキーワードどうしの関連付けを含む複数の要因に基づいて算出する類似度付関連付点数算出部である。

【0173】以上のように構成されたシステムが実行されるハードウェア構成は図2に示した第1の発明の実施の一形態のハードウェア構成と同じである。したがって、詳細な説明は省略する。

【0174】以上のように構成された情報要約方法、および情報要約装置の動作を図21のフローチャートを参照しながら説明する。

【0175】ステップg1～ステップg3では、第5の発明の実施の一形態のステップe1～ステップe3と同様の処理を行う。

【0176】ステップg4では、第3の発明の実施の一形態のステップc4と同様の処理を行う。ただし、ステップc4では、2つのキーワードのうち片方が、もう一方の部分になっている場合には同一と見なす処理を行ったが、本実施の一形態では完全に一致するキーワード以外は同一とは見なさない。なお、本実施の一形態においても、2つのキーワードのうち片方がもう一方の部分になっているキーワードを同一と見なす処理を施しても構わない。

【0177】ステップg5では、ステップg3で切り出したキーワードに対して、第5の発明の実施の一形態のステップe4と同様の処理を行う。また、ステップg4で関連付けられたキーワードの関連付けに対しては、完全に一致する関連付けをまとめ、関連付けの出現頻度を抽出された個数にする。たとえば、(泉佐野、ヘリ墜落)が4個切り出された場合、関連付け(泉佐野、ヘリ墜落)を1つにまとめて出現頻度を4にする。また、得られたキーワード、および関連付けのカウント結果はFILO(First In Last Out)のスタックに単位を識別する識別子と組み合わせ、キーワードから先に記憶(プッシュ)する。単位を識別する識別子と組み合わせる方法、および理由は第5の発明の実施の一形態のステップe4と同様である。関連付けから先に記憶するのは、後述のステップg6でこのスタックからデータを読み出す際に、関連付けを先に取り出し、その後にキーワードを取り出し、キーワードの類似性を算出する際に先に取り出した関連付けを用いるためである。

【0178】なお、このステップにおいて同一ではない

67

キーワードを含む 2 つの関連付け (K 1, K 2) (L 1, L 2) に対しても、関連付けどうしの類似度を算出し、その類似度に基づいて出現頻度を加算してもよい。関連付けどうしの類似度は、K 1 と L 1 の類似度、K 2 と L 2 の類似度を第 5 の発明の実施の一形態のステップ e4 と同様の方法で算出し、この算出結果の平均値にする。また、K 1 と L 2 の類似度、K 2 と L 1 の類似度を算出して平均した値の方が大きければ、その値を関連付けどうしの類似度とする。たとえば、関連付け (米朝高官会議、核問題) と (米朝高官協議、核問題討議) の類似度は、「米朝高官会議」と「米朝高官協議」の類似度 0. 8 3 と「核問題」と「核問題討議」の類似度 0. 7 5 の平均を求めて、 $(0. 8 3 + 0. 7 5) / 2 = 0. 7 9$ (小数点 3 位以下四捨五入) になる。関連付けどうしの類似度が算出されれば、あとはキーワードに対してステップ e4 で施した処理を関連付けに対しても行う。2 つの関連付け A, B の間で算出した類似度を S として、A の出現頻度に $\{S \times (B \text{ の出現頻度})\}$ を加え、B の出現頻度に $\{S \times (A \text{ の出現頻度})\}$ を加える。たとえば、類似度を算出する前の関連付け (米朝高官会議、核問題) の出現頻度が 3 で、(米朝高官協議、核問題討議) の出現頻度が 5 であった場合、類似度 0. 7 9 を用いて (米朝高官会議、核問題) の出現頻度は $3 + 5 \times 0. 7 9 = 6. 9 5$ になり、(米朝高官協議、核問題討議) の出現頻度は $5 + 3 \times 0. 7 9 = 7. 3 7$ になる。このような関連付けの類似度は、ステップ e3 で切り出されたキーワードすべての組み合わせに対して算出する。ただし、計算時間を少なくするためにキーワードで一致する文字数が 5 割以上のもののみに対して類似度を算出するなどの方法も考えられる。

30

68

【0 1 7 9】ステップ g6 からステップ g8 では、第 5 の発明の実施の一形態のステップ e5 からステップ e7 と同様の処理を行う。ただし、第 5 の発明ではキーワードに対して行っていた処理を関連付けに対しても行う点が異なる。

【0 1 8 0】ステップ g9 では、ステップ g6 で取り出したデータがキーワードである場合と、関連付けである場合で処理が異なる。取り出したデータが関連付けである場合、第 5 の発明の実施の一形態のステップ e8 と同様の処理を行う。ただし、ステップ e8 では、キーワード間での類似度を算出して単位内頻度と単位間共有度を算出したのに対し、このステップでは関連付けどうしの類似度を算出し、関連付けの単位内頻度と単位間共有度を算出、更新する。関連付けどうしの類似度はステップ g5 の説明の中で述べた通りである。

【0 1 8 1】ステップ g6 で取り出したデータがキーワードである場合、キーワードの類似度のみでなく、キーワードと関連付けされているキーワードの類似度も考慮したマクロ類似度を用いて単位間共有度を算出、更新する。単位内頻度については、第 5 の発明の実施の一形態のステップ e8 と同様の方法で更新する。すなわち、キーワード間の類似度に基づいて更新する。単位間共有度については、まずステップ g6 で取り出したキーワードを含む関連付けをステップ g11 で記憶されたデータから取り出して集合 I とする。次に、キーワード頻度表のキーワードに対し、そのキーワードを含む関連付けの集合 J をキーワード頻度表から取り出す。たとえば、以下の (表 1 9) に示すキーワード頻度表を想定する。

【0 1 8 2】

【表 1 9】

キーワード、あるいは関連付け	単位内頻度	単位間共有度
核問題	F i 1	F e 1
調印式	F i 2	F e 2
米朝高官協議	F i 3	F e 3
(核問題, 調印式)	F i 4	F e 4
(調印式, 米朝高官協議)	F i 5	F e 5
(核問題, 米朝高官協議)	F i 6	F e 6
ヨルダン	F i 7	F e 7
仮調印式	F i 8	F e 8
イスラエル	F i 9	F e 9
(仮調印式, ヨルダン)	F i 10	F e 10
(仮調印式, イスラエル)	F i 11	F e 11
(ヨルダン, イスラエル)	F i 12	F e 12
:	:	:
:	:	:

【0183】ステップg6で取り出されたデータのキーワードが「調印式典」とあるとする。このデータが取り出される前に、ステップg11で記憶されたデータが以下の*30 * (表20) であると想定する。
【0184】
【表20】

キーワード、あるいは関連付け	出現頻度	単位識別子
(米朝高官会談、調印式典)	F 1	Bid1
(核問題、調印式典)	F 2	Bid1
北朝鮮	F 3	Bid1

【0185】集合Iは、{(米朝高官会談, 調印式典), F 1, Bid1}, {(核問題, 調印式典), F 2, Bid1} である。

【0186】取り出されたキーワードとキーワード頻度表内のキーワードとの類似度の算出と頻度の更新処理は、キーワード頻度表内の各キーワードに対して順に行う。このため、ここではキーワード頻度表内の「調印式」との間での類似度計算について述べる。集合Jは{(核問題, 調印式), F i 4, F e 4}, {(調印式, 米朝高官協議), F i 5, F e 5} になる。

【0187】集合I, Jが求められると、2つの集合の間で、類似度計算の対象となっているキーワードと関連付けられたキーワードを取り出し、それらのキーワード間で類似度を算出する。ただし、キーワードの集合どうしであるために、集合Iから一つずつキーワードを取り出し、取り出したキーワードと集合Jの全てのキーワードとの間で順に類似度を算出し、最大の類似度を集合Iから取り出したキーワードに対する類似度とする。このようにして、集合Iに属する各キーワードに対する類似度を求め、その平均を集合Iと集合Jの集合類似度Kと

する。たとえば、上記の例では集合Iからは〔米朝高官会談、核問題〕、集合Jからは〔核問題、米朝高官協議〕が取り出されて、集合Iの「米朝高官会談」に対して集合Jの「米朝高官協議」が最大の類似度となり、類似度は0.67、「核問題」に対しては集合Jの「核問題」が最大の類似度となり、その類似度は1.0である。これらの類似度の平均を求めると0.84になる。

(小数点以下第3位を四捨五入)したがって、集合Iと集合Jの集合類似度Kは0.84である。この集合類似度Kと、キーワード「調印式」と「調印式典」の類似度

S=0.86との、重み付き和をキーワード間のマクロ類似度とする。

【0188】重みは、あらかじめ定めた定数X、Yであり、マクロ類似度は $X \times K + Y \times S$ になる。マクロ類似度が求められれば、あとは第5の発明の実施の一形態のステップe8と同様の処理でキーワード頻度表内の単位内頻度と単位間共有度を更新する。ただし、ステップe8と比べて、キーワードの類似度の代わりにマクロ類似度を用いる。これにより、よく似たキーワードが別の話題で扱われている場合に、マクロ類似度が低くなり、同じ話題で扱われている場合に、マクロ類似度が大きくなる。たとえば、ステップg6で取り出されたキーワードが同じ「調印式典」であっても、キーワード頻度表内の比較対象が「仮調印式」である場合、集合Jが〔ヨルダン、イスラエル〕になる。この結果、「調印式典」と「仮調印式」の類似度は0.75と高いが、集合Iと集合Jの類似度は0になり、「調印式典」と「仮調印式」のマクロ類似度は小さな値になる。これにより、同じ調印式に関することであっても、別の話題の調印式であり、マクロ類似度がそのことを表している。

【0189】なお、本実施の一形態では集合I、Jに含まれるキーワードの出現頻度、単位内頻度、単位間共有度を使わずに集合類似度Kを算出したが、算出の際にこれらの値を含めても構わない。これらの値を含めれば算出方法が複雑になるが、キーワード間の関連付けの強さ(どれだけ多くの回数関連付けられているか)を考慮した上でマクロ類似度を算出することができる。

【0190】ステップg10～ステップg14では、第5の発明の実施の一形態のステップe9～ステップe13と同様の処理を行う。ただし、第5の発明ではキーワードに対して行っていた処理を関連付けに対しても行う点が異なる。

【0191】ステップg15では、キーワード、および関連付けの出現頻度、単位内頻度、単位間共有度の値が整数ではなく、小数点以下のある実数になっている点異なるが、第3の発明の実施の一形態のステップc14と同様の処理を行う。

【0192】なお、本発明においても第3、第4の発明と同様に、キーワードどうしの関連付けのみを用いたり、単位間共有度のみを用いる方法が考えられる。特

に、キーワードを用いなくてキーワードどうしの関連付けのみで、かつ単位内頻度を用いなくて単位間共有度のみを用いる方法について本発明の上記実施の一形態で説明したステップのうち異なる部分についてのみ説明する。

【0193】まず、ステップg5においてキーワードに対する処理を行わずにキーワードどうしの関連付けに対する処理のみを行う。また、関連付けどうしに対する処理のうち、類似度の算出処理、出現頻度に関する処理も行わない。すなわち、キーワードについてはスタックに記憶せずに、キーワードどうしの関連付けのみをスタックに記憶する。その際には、出現頻度に関する処理を行わないために、カウント表の形式は第3の発明で説明した(表15)のような形式になる。

【0194】また、ステップg6、ステップg8では、ステップg6で取り出されるデータにはキーワードはなく、キーワードどうしの関連付けに関するデータのみになる点異なるが基本的な処理内容は変わらない。

【0195】ステップg9では、ステップg6で取り出されるデータがキーワードどうしの関連付けのみになることから、ステップg6で取り出されたデータがキーワードである場合に行っていた処理は不要になる。このため、キーワードどうしの関連付けに対して行っていた処理のみを行う。この結果、キーワード間の類似度の算出に用いていたマクロ類似度の算出処理も不要になる。また、第3の発明の実施の一形態において説明したのと同様に、単位間共有度のみを用いて要約を行うために、キーワード頻度表は(表16)の形式になり、ステップg9での処理においては、単位間共有度に対する処理のみを行い、単位内頻度に対する処理は省略する。

【0196】ステップg15においては単位間共有度の値が実数になっている点異なるが第3の発明の実施の一形態で説明した(表16)に対する処理と同様にしてキーワードを選択することが出来る。

【0197】次に、第8の発明にかかる実施の一形態を図面を参照しながら説明する。

【0198】第8の発明の実施の一形態として情報要約付き文字放送受信装置を示す。図22は、第8の発明の実施の一形態のシステム構成図である。図22には第4の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図22において、81は文字放送キーワード切り出し部23で切り出したキーワードに対して類似性を考慮して点数を算出する文字放送類似度付関連付け点数算出部である。

【0199】以上のように構成されたシステムが実行されるハードウェア構成は図8に示した第2の発明の実施の一形態のハードウェア構成と同じである。したがって、詳細な説明は省略する。

【0200】以上のように構成されたキーワード間の類

似性を考慮した情報要約機能付き文字放送受信装置の動作を図 2 3 のフローチャートを参照しながら説明する。

【0 2 0 1】ステップh1～ステップh2では、第 2 の発明の実施の一形態のステップb1～ステップb2と同様の処理を行う。

【0 2 0 2】ステップh3～ステップh14は、文字放送類似度付関連付け点数算出部 8 1 で実行する。

【0 2 0 3】ステップh3～ステップh14は、第 7 の発明の実施の一形態のステップg3～ステップg14と同様の処理を行う。ただし、キーワード頻度表の項目名が異なり、単位内頻度が番組内頻度に、単位間共有度が番組間共有度になっている。内容的には同じであり、単位内頻度に対して行った処理と同じ処理を番組内頻度に対して行い、単位間共有度に対して行った処理と同じ処理を番組間共有度に対して行う。ステップh15では、第 7 の発明の実施の一形態のステップg15と同様の処理を行う。ただし、出力部 2 に出力する代わりに表示部 2 6 に出力する。

【0 2 0 4】なお、本発明においても第 7 の発明と同様に、単位間共有度のみを用いて選択する方法、キーワードどうしの関連付けのみを用いる方法が考えられる。この方法は、基本的に第 7 の発明のなお書きの部分で説明したものと同一処理で実現できるために、詳細な説明は省略する。

【0 2 0 5】次に、第 9 の発明にかかる実施の一形態を図面を参照しながら説明する。

【0 2 0 6】第 9 の発明の実施の一形態としてキーワードとして用いることのできない例外のキーワードを考慮した情報要約方法、および情報要約装置を示す。図 2 4 は、第 9 の発明の実施の一形態のシステム構成図である。図 2 4 には第 1 の発明の実施の一形態の一構成例と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図 2 4 において、9 1 はキーワードとして処理しないキーワードを記憶しておく例外キーワード記憶部である。9 2 は、例外キーワード記憶部 9 1 に記憶されているキーワードに基づいてキーワードを抽出する例外付キーワード切り出し部である。

【0 2 0 7】以上のように構成されたシステムが実行されるハードウェア構成は図 2 5 に示す。図 2 5 に示したハードウェア構成の各部分は第 1 の発明の実施の一形態のハードウェア構成の部分と図 2 4 に示したシステム構成の構成部分からなるので、同一構成部分に同一番号を付して詳細な説明は省略する。

【0 2 0 8】以上のように構成された情報要約方法、および情報要約装置の動作を表すフローチャートは、図 3 に示した第 1 の発明の実施の一形態の動作を表すフローチャートのステップa3とステップa4の間に図 2 6 のフローチャートに示すステップを挿入したものである。したがって、図 2 6 に示すステップについてのみ説明する。

また、ステップa3は例外付キーワード切り出し部 9 2 で実行される。

【0 2 0 9】ステップi1は、ステップが実行される度にステップa3で切り出したキーワードを一つずつ順に読み出す。このステップが何回も呼び出されると、ステップa3で切り出したキーワードを順にすべて読み出してしまい、読み出すべきキーワードが無くなった時点で図 2 6 のフローチャートの”END”の方へ進む。このステップでは、順に読み出すだけで、読み出されたデータはこのステップによって消去されない。ステップi1は、例外付キーワード切り出し部 9 2 で実行される。

【0 2 1 0】ステップi2は、例外キーワード記憶部 9 1 に記憶されているキーワードからキーワード 1 つを読み出す。例外キーワード記憶部 9 1 には、以下の(表 2 1)に示すように、キーワードとして処理しないキーワードが予め記憶されている。

【0 2 1 1】

【表 2 1】

例外キーワード
a
the
into
according
:

【0 2 1 2】例外キーワード記憶部 9 1 からのキーワードの読み出しは、このステップが実行される度に次のキーワードへと順に進める。たとえば、初めてこのステップが実行されたときには(表 2 1)からは、”a”が読み出され、次に実行されたときには”the”が読み出される。このように、順に読み出していくと実行が繰り返されるうちに読み出すべきデータが無くなる。その場合には、図 2 6 のフローチャート内の”END”で示される方へ進む。

【0 2 1 3】ステップi3は、ステップi1で読み出したキーワードとステップi2で読み出したキーワードを比較する。比較の結果、等しい場合にはステップi4へ進み、その他の場合にはステップi2へ戻る。

【0 2 1 4】ステップi4は、ステップa3で切り出したキーワードの中からステップi1で読み出したキーワードを消去する。これにより、ステップa4以降では消去されたキーワードは処理対象にならない。

【0 2 1 5】次に、第 9 の発明の第 2 の実施の一形態として、図 1 0 に示す第 3 の発明の実施の一形態のシステム構成のうち、キーワード切り出し部 3 を図 2 9 に示す構成で置き換えたシステム構成のものがある。入力部 1

が図 29 の端点 1 に接続され、図 29 の端点 2 がキーワード関連付部 31 に接続される。この実施の一形態における動作は、図 11 に示すフローチャートのステップ c3 とステップ c4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 25 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0216】次に、第 9 の発明の第 3 の実施の一形態として、図 16 に示す第 5 の発明の実施の一形態のシステム構成のうち、キーワード切り出し部 3 を図 29 に示す構成で置き換えたシステム構成のものがある。入力部 1 が図 29 の端点 1 に接続され、図 29 の端点 2 が類似度付点数算出部 51 に接続される。この実施の一形態における動作は、図 17 に示すフローチャートのステップ e3 とステップ e4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 25 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0217】次に、第 9 の発明の第 4 の実施の一形態として、図 20 に示す第 7 の発明の実施の一形態のシステム構成のうち、キーワード切り出し部 3 を図 29 に示す構成で置き換えたシステム構成のものがある。入力部 1 が図 29 の端点 1 に接続され、図 29 の端点 2 がキーワード関連付け部 31 に接続される。この実施の一形態における動作は、図 21 に示すフローチャートのステップ g3 とステップ g4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 25 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0218】次に、第 10 の発明にかかる実施の一形態を図面を参照しながら説明する。

【0219】第 10 の発明の実施の一形態としてキーワードとして用いることのできない例外のキーワードを考慮した文字放送受信装置を示す。図 27 は、第 10 の発明の実施の一形態のシステム構成図である。図 27 には第 2 の発明の実施の一形態の一構成例と、第 9 の発明の実施の一形態と同一の構成部分があるので、同一構成部分に対しては同一番号を付して詳細な説明は省略する。図 27 において、101 は文字放送として受信した番組に対して、例外キーワード記憶部 91 に記憶されているキーワードに基づいてキーワードを抽出する例外付文字放送キーワード切り出し部である。

【0220】以上のように構成されたシステムが実行されるハードウェア構成は図 28 に示す。図 28 に示したハードウェア構成の各部分は第 2 の発明の実施の一形態のハードウェア構成の部分と図 27 に示したシステム構成の構成部分からなるので、同一構成部分に同一番号を付して詳細な説明は省略する。

【0221】以上のように構成された情報要約方法、および情報要約装置の動作を表すフローチャートは、図 9

に示した第 2 の発明の実施の一形態の動作を表すフローチャートのステップ b3 とステップ b4 の間に図 26 のフローチャートに示すステップを挿入したものである。したがって、すべてのステップについて説明済みであり詳細な説明は省略する。ただし、ステップ b3、ステップ i1 ～ステップ i4 は例外付文字放送キーワード切り出し部 101 で実行される。

【0222】次に、第 10 の発明の第 2 の実施の一形態として、図 13 に示す第 4 の発明の実施の一形態のシステム構成のうち、文字放送キーワード切り出し部 23 を図 30 に示す構成で置き換えたシステム構成のものがある。文字放送受信部 21 が図 30 の端点 1 に接続され、図 30 の端点 2 が文字放送キーワード関連付け部 41 に接続される。この実施の一形態における動作は、図 14 に示すフローチャートのステップ d3 とステップ d4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 28 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0223】次に、第 10 の発明の第 3 の実施の一形態として、図 18 に示す第 6 の発明の実施の一形態のシステム構成のうち、文字放送キーワード切り出し部 23 を図 30 に示す構成で置き換えたシステム構成のものがある。文字放送受信部 21 が図 30 の端点 1 に接続され、図 30 の端点 2 が文字放送類似度付点数算出部 61 に接続される。この実施の一形態における動作は、図 19 に示すフローチャートのステップ f3 とステップ f4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 28 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0224】次に、第 10 の発明の第 4 の実施の一形態として、図 22 に示す第 8 の発明の実施の一形態のシステム構成のうち、文字放送キーワード切り出し部 23 を図 30 に示す構成で置き換えたシステム構成のものがある。文字放送受信部 21 が図 30 の端点 1 に接続され、図 30 の端点 2 が文字放送キーワード関連付け部 41 に接続される。この実施の一形態における動作は、図 23 に示すフローチャートのステップ h3 とステップ h4 の間に図 26 のステップを挿入したものであり、各ステップはすでに説明済みなので詳細な説明は省略する。また、ハードウェア構成は図 28 に示す第 1 の実施の一形態のハードウェア構成と同じである。

【0225】なお、第 1 の発明、第 3 の発明、第 5 の発明、第 7 の発明、第 9 の発明のいずれの発明も、オーディオ装置、電子メール装置、パーソナルコンピュータへの適用が考えられる。オーディオ装置への適用においては、ラジオの文字放送を対象としたものであり、単位としては放送局ごとの放送内容になる。また、段落は放送内容の中の固まりであり、たとえばニュース放送では記

事の切れ目が段落の切れ目になる。オーディオ装置の中には、たとえば車載用オーディオがあり音楽を聴きながら文字放送で提供されるニュース情報の要約のみを見ろといった使い方ができる。従来のオーディオの場合、ニュースを知りたければ音楽番組などを諦めてニュース番組を聴かなければならなかったのに対し、音楽番組も、またニュースについても同時に知ることができる。さらに、オーディオ機器はテレビに比べて大画面を備えることが困難であり、文字放送として送られてくるニュースは短い文章で構成されているものの、ニュースの数が多ければ何ページものニュース番組になる。この結果、要約してキーワードの組み合わせといった簡潔な形式で表示することによる効果は非常に大きい。

【0226】電子メールでは、多数の人が文書を投稿する伝言板のような機能があり、電子メール装置に適用する場合には、単位としては利用者ごとの文書になる。また、段落としては、文書の中の段落を考えてもよいが、1文書1段落ととらえることもできる。現在でも、インターネットにおいてニュースグループとして特定分野の研究者が互いの情報交換を行っているなど、伝言板的に電子メールを利用することは、電子メールの利点の一つである。このような、ニュースグループの情報に対して、本発明を適用することにより、特定の分野でどのような話題が注目されているのかがキーワードの組み合わせといった簡単な形式でわかる。今後、パソコン通信などで電子メール機能が普及すれば、このようなニュースグループの利用者数と比べて爆発的に人数が増え、ニュースグループのような伝言板的機能へのメールの投稿量が増加する。その結果、すべての情報にいちいち目を通すだけで長時間を費やさなければならなくなり、本発明の要約機能は極めて効果的に働く。

【0227】パーソナルコンピュータへ適用した場合、ワープロなどで作成した文書が1単位になる。また、段落としては、文書の中の段落を考えてもよいが、1文書1段落ととらえることもできる。また、パーソナルコンピュータに文字放送受信機能を備えることも可能であり、その場合には第2の発明、第4の発明、第6の発明、第8の発明と同様に考えることができる。ワープロで作成した文書の中には、たとえば学会の論文集をテキスト情報として蓄えたものが考えられる。このような、テキスト情報に対して、本発明を適用すれば、学会での主な話題をキーワードの組み合わせといった簡潔な表現で把握する事ができる。また、特許情報などが提供されており、本発明を適用すれば時代と共に特許としてどのようなものが多数出願されているかが簡単に把握できる。

【0228】なお、第1、3、5または7の発明の実施の他の形態として、更に、次に示すような、例が考えられる。

【0229】即ち、利用者が表示させたいと希望するデ

ータを指定出来るデータ指定手段を新たに設ける。入力手段から入力されて、キーワードを抽出するのに利用したデータを、一旦、作業領域に記憶しておく。

【0230】上述した内容と同様に、出力手段に情報要約を出力した後に、データ指定手段からの入力を受け付け、この入力に基づいて作業領域に記憶しておいたデータを検索し、その検索されたデータを利用者に表示することも考えられる。

【0231】この場合のシステム構成のうち新たに設ける部分を図31に示す。図31において、111は、利用者から、利用者が表示させたいと希望するデータを指定するための指定入力を受け付けるデータ指定手段である。即ち、利用者は、出力部2に表示された情報要約を見て、その中から、更に詳しい内容を表示させたいと思ったキーワードを、データ指定手段111を用いて指定する。112は、データ記憶部である。113は、利用者からの上記指定入力に基づいて、作業領域としてのデータ記憶部112に一旦記憶されたデータを、検索する検索部である。114は、検索部113で検索したデータを出力するデータ提示部である。

【0232】この構成図で示される情報要約方法、および情報要約装置の実行されるハードウェア構成は、基本的に第1、3、5または7の発明の実施の一形態において示したハードウェア構成にデータ指定手段が付加されたものである。また、フローチャートは、基本的に第1、3、5または7の発明の実施の一形態において示したフローチャートに、それぞれステップa1、ステップc1、ステップe1、ステップg1のデータ入力処理の後に主記憶装置12の作業領域上にデータを記憶するデータ記憶処理を加え、さらにそれぞれステップa10、ステップc14、ステップe14、ステップg15で示される要約出力処理の後に、データ指定手段111からの指定入力を受け付ける処理、受け付けた指定入力に基づいてデータを検索する処理、検索結果を表示する処理を加えたものである。

【0233】データ記憶処理は、作業領域上に入力データをそのままの形式で記憶する。指定入力を受け付ける処理は、図6や図12に示したような出力に対して、マウス、十字キー、数字ボタンなどを用いて利用者がキーワードを選択するのを受け付ける。これらの選択の方法は、従来より用いられ得ている一般的な方法で構わない。受け付けた指定入力に基づいてデータを検索する処理は、利用者が選択したキーワードを含むデータの部分をデータ記憶処理で記憶したデータから検索する。この時、利用者が選択したキーワードを含む段落、行、前後5行など予め定めた量のデータ部分を検索結果として取り出す。処理結果を表示する処理は、検索されたデータ部分を出力部2に表示する。たとえば、利用者がキーワード「朝日新聞社」を選択した場合、図5に示したデータ例の中からは、単位1の第1段落、単位3の第2段

落、が検索されて出力される。ただし、図 5 の各単位は空白行で段落が分かれているものとする。

【0234】また、このようなデータ表示機能は、第 9 の発明とも組み合わせることが出来る。

【0235】なお、第 2、4、6 または 8 の発明の実施の他の形態として、更に、次に示すような、例が考えられる。

【0236】即ち、利用者が表示させたいと希望するデータを指定出来るデータ指定手段を新たに設ける。文字放送受信手段から受信して、キーワードを抽出するの

10 利用したデータを、一旦、作業領域に記憶しておく。
【0237】上述した内容と同様にして、表示手段に情報要約を表示した後に、データ指定手段からの入力を受け付け、この入力に基づいて作業領域に記憶しておいたデータを検索し、その検索されたデータを利用者に表示することも考えられる。

【0238】この場合のシステム構成のうち新たに設ける部分を図 3 2 に示す。図 3 2 の構成部分は、図 3 1 の構成部分と同一構成部分からなるので、その部分に同一番号を付して説明は省略する。この構成図で示される文字放送受信装置の実行されるハードウェア構成は、基本的

20 的に第 2、4、6 または 8 の発明の実施の一形態において示したハードウェア構成にデータ指定手段が付加されたものである。また、フローチャートは、基本的に第 2、4、6 または 8 の発明の実施の一形態において示したフローチャートに、それぞれステップ b2、ステップ d2、ステップ f2、ステップ h2 の番組受信処理の後に主記憶装置 1 2 の作業領域上に受信した番組のデータを記憶するデータ記憶処理を加え、さらにそれぞれステップ b10、ステップ d14、ステップ f14、ステップ g15 で示される

30 要約出力処理の後に、データ指定手段 1 1 1 からの指定入力を受け付ける処理、受け付けた指定入力に基づいてデータを検索する処理、検索結果を表示する処理を加えたものである。各処理については、上述の第 1、3、5 または 7 の発明とデータ表示機能を組み合わせた場合と同じであり、説明を省略する。

【0239】また、このようなデータ表示機能は、第 10 の発明とも組み合わせることが出来る。

【0240】なお、第 1、3、5 または 7 の発明の実施の他の形態として、更に、次に示すような、例が考え

40 られる。
【0241】即ち、利用者が興味をもっているデータ内容を、キーワードを用いて指定出来るデータ指定手段を新たに設ける。

【0242】上述した内容と同様にして、出力手段に要約結果を出力した後に、このデータ指定手段からキーワードの指定入力を受け付け、このキーワードの指定入力があり、かつ入力部 1 から再びデータの入力があった場合に、前記指定入力されたキーワードが入力データ中に存在するか否かを検索し、検索の結果前記指定入力され

たキーワードが含まれている場合に出力部 2 に前記キーワードを含むデータの全部又は一部を表示することが考えられる。あるいは、表示内容としては、データ内容そのものではなく、前記キーワードを含むデータが存在したことを知らせるメッセージを表示することも考えられる。これらデータ内容あるいは、メッセージの表示は、上述のようにキーワードの指定入力があり、かつ入力部 1 から再びデータの入力があった場合、更に、その入力されたデータ内容が、以前に入力された内容から更新されているか否かを判定して、更新されている場合のみ、行なうものである。

【0243】尚、利用者がデータ指定手段 1 1 1 を用いて指定するキーワードは、出力部 2 に表示された情報要約の中から、指定してもよいし、あるいは、情報要約の中には無い、任意のキーワードを指定してもよい。

【0244】又、データ内容が更新されているか否かの判定は、図 3 4、3 5 を用いて後述するように、入力されたデータ内に、指定されたキーワードが含まれているか否かを検索するステップの前に行なってもよいし（図 3 5 参照）、あるいは、後で行なってもよい（図 3 4 参照）。

【0245】この場合のシステム構成のうち新たに設ける部分を図 3 3 に示す。図 3 3 において、図 3 1 と同一構成部分を含むので、同一部分には同一番号を付し説明は省略する。1 2 1 は、入力部 1 から再びデータが入力された時に、データ指定手段 1 1 1 で受け付けたキーワードの指定入力に基づいてデータを検索するキーワード発見部、1 2 2 はキーワード発見部 1 2 1 でキーワードを含むデータが発見されたときに利用者に知らせる連絡部である。この構成図で示される情報要約方法、および情報要約装置の実行されるハードウェア構成は、基本的に第 1、3、5 または 7 の発明の実施の一形態において示したハードウェア構成にデータ指定手段が付加されたものである。

【0246】また、フローチャートは、基本的に第 1、3、5 または 7 の発明の実施の一形態において示したフローチャートに、それぞれステップ a10、ステップ c14、ステップ e14、ステップ g15 で示される要約出力処理の後に、図 3 4、図 3 5 に示す処理を加えたものである。以下、図 3 5 を中心に、各ステップについて説明する。

【0247】ステップ j1 は、データ指定手段 1 1 1 において、利用者から興味のあるキーワードの指定を受け付け、指定されたキーワードを記憶する。指定入力を受け付ける処理は、図 6 や図 1 2 に示したような出力に対して、マウス、十字キー、数字ボタンなどを用いて利用者がキーワードを選択するのを受け付ける。これらの選択の方法は、従来より用いられ得ている一般的な方法で構わない。

【0248】ステップ j2 では、ステップ j1 で指定入力があったか否かを判定する。指定入力があればステップ j3

へ進み、無い場合には処理を終了する。

【0249】ステップj3では、キーワード発見部121において、入力部1から新たなデータの入力があるか否かを判定する。尚、放送のように常にデータが入力されている場合には、このステップでは前回入力されたデータとの比較を行わずに、データが更新されたか否かの比較・判定は、後述するステップj5において行なう。メールのように、一固まりのデータとして入力される場合には新たなデータの到着を検出することにより判定できる。

【0250】ステップj4～ステップj6では、キーワード発見部121において、ステップj1で指定入力され、記憶されたキーワードが、新たに入力され、且つ更新されたデータに含まれるか否かを検索し、含まれている場合にはステップj7へ進み、含まれていない場合にはステップj3へ戻る。この検索の方法は、単純にキーワードが含まれているかどうかを調べる従来のキーワードマッチングの手法でも構わない。

【0251】ステップj7では、ステップj4での処理の結果、入力データ中に指定入力されたキーワードが含まれており、そのデータ部分を出力部2に表示する。たとえば、利用者がキーワード「朝日新聞社」をステップj1で指定入力しており、図5に示したデータ例が新たに入力された場合、単位1の第1段落、単位3の第2段落、にキーワードが含まれており、これらを出力部2に出力する。ただし、図5の各単位は空白行で段落が分かれているものとする。また、この処理は図36に示すようにキーワードを含むデータが入力されたことだけをメッセージとして表示し、利用者が別に設けたボタンを押した場合に、キーワードを含むデータ部分を表示するなどの方法を用いても構わない。さらに、この処理においてデータを表示するのみでなく、キーワードを含むデータを蓄積しておくことも考えられる。これにより、興味のあるキーワードを含むデータ部分が新たに入力されたデータに含まれれば、次々に蓄積されるといった効果が得られる。

【0252】また、以上のような処理は、第9の発明とも組み合わせることが出来る。

【0253】なお、第2、4、6または8の発明において、利用者から興味のあるデータの指定入力を受け付けるデータ指定手段を新たに設け、表示手段に要約結果を出力した後に、このデータ指定手段からキーワードの指定入力を受け付け、このキーワードの指定入力があり、かつ文字放送受信部21で受信したデータが更新された場合に、前記指定入力されたキーワードが受信データ中に存在するか否かを検索し、検索の結果前記指定入力されたキーワードが含まれている場合に表示部26に前記キーワードを含むデータ部分、もしくは前記キーワードを含むデータが存在したことを知らせるメッセージを表示することも考えられる。この場合のシステム構成のうち新たに設ける部分は図33に示したシステム構成と殆

ど同様であり、異なる点はキーワード発見部は入力部1ではなく、文字放送受信部21と接続され、連絡部122は出力部2ではなく、表示部26と接続される点である。この文字放送受信装置の実行されるハードウェア構成は、基本的に第2、4、6または8の発明の実施の一形態において示したハードウェア構成にデータ指定手段が付加されたものである。また、フローチャートは、基本的に第2、4、6または8の発明の実施の一形態において示したフローチャートに、それぞれステップb10、ステップd14、ステップf14、ステップh15で示される要約出力処理の後に、図35に示す処理を加えたものである。ただし、図35のステップのうち、ステップj5では、文字放送受信部21から受信されたデータを、前回受信されたデータと比較し、データが更新されたか否かを判定する。

【0254】また、ステップj7では、ステップj4での処理の結果、指定入力されたキーワードが更新されたデータに含まれており、そのデータ部分を表示部26に表示する。たとえば、利用者がキーワード「朝日新聞社」をステップj1で指定入力しており、図5に示したデータが新たに更新されたデータとして受信された場合、単位1の第1段落、単位3の第2段落、にキーワードが含まれており、これらを表示部26に表示する。また、この処理は図36に示すようにキーワードを含むデータが入力されたことだけをメッセージとして表示し、利用者が別に設けたボタンを押した場合に、キーワードを含むデータ部分を表示するなどの方法を用いても構わない。

【0255】尚、図34は、データ内容が更新されているか否かの判定を実行する順番は、上述した図35に示す内容と異なり、入力されたデータ内に指定されたキーワードが含まれているか否かを検索するステップj6の後で行なった場合のフローチャートである。図34でのステップの番号は、図35と同じ内容のものには、同じ番号を付した。

【0256】また、以上のような処理は、第10の発明とも組み合わせることも出来る。

【0257】以上述べたところより明かなように、第1の発明にかかる実施の形態によれば、所定の単位で与えられる文字列データが多数ある場合に、それらのデータ間に共通して現れ、かつ頻度が大きいキーワードが共通の話題として抽出される。この結果、本発明による情報要約装置の利用者にとっては多数のデータをすべて見ることなく短時間で主だった話題を把握することができるといった利点が生じる。

【0258】第2の発明にかかる実施の形態によれば、文字放送といったリアルタイムに与えられるデータに対する情報要約を行うことにより、常に最新データの主だった話題を抽出し、画面上に情報要約として表示することができる。このため、たとえばニュース放送を文字放送で受信すれば、最新の主だったニュースを常に要約と

して把握することができる。この結果、膨大な量のニュース放送を見ることなく主な内容だけを常に要約として知ることができる。

【0259】第3の発明にかかる実施の形態によれば、所定の単位で与えられる多数の文字列データに対し、その主な内容を複数のキーワードで表示する際に、キーワード間の関連を明らかにして表示することができる。この結果、単純なキーワードの羅列と異なり、キーワード間の関連に基づき、話題の内容がより解りやすくなる。

【0260】第4の発明にかかる実施の形態によれば、文字放送といったリアルタイムに与えられるデータから常に最新データの主だった話題を抽出し、画面上に情報要約として表示する際に、複数のキーワードをキーワード間の関連を明らかにした形式で表示する。このため、最新ニュースで主に取り上げられている話題を互いに関連のあるキーワードの集合として把握することができる。この結果、膨大な量のニュース放送を見ることなく主な内容だけを短時間に、かつ解りやすい形の要約として知ることができる。

【0261】第5の発明にかかる実施の形態によれば、キーワードの頻度を扱う場合に、同じことを表していて表現の異なるキーワードを処理することができる。たとえば、同じ選手の名前でも「伊達公子」と表現したり、「伊達選手」と表現したりする。このように表現が異なっても同じ意味で用いられているキーワードは、類似度を計算することにより類似キーワードとして互いの頻度を加算する。また、頻度の加算の際には、類似度を掛け合わせるために、より類似度の大きいキーワードが抽出されれば、その類似度に合わせてより大きな値が頻度として加算される。類似度の小さいキーワードが抽出されても、類似度を掛け合わせるために小さい値が頻度として加算される。この結果、内容的に有用なキーワードであるにもかかわらず別表現されていることが原因で頻度が小さくなり、要約結果として用いられなくなるといったことがなくなり、より適切な要約結果を得ることができる。

【0262】第6の発明にかかる実施の形態によれば、文字放送といった複数の放送局からデータが提供され、データを作成した作成者が異なる場合において別表現されたキーワードの処理が可能である。特に、放送内容がニュースなど世の中の出来事を扱っている場合、専門用語がきちんと決まっていないという問題があり、キーワードの別表現は頻繁に起こる。したがって、別表現されたキーワードでも類似キーワードとして頻度を加算することにより、多くの放送局で共通に扱われている話題に関する有用なキーワードを要約結果に反映させることができる。

【0263】第7の発明にかかる実施の形態によれば、表現の異なるキーワードの類似性を算出する場合、2つのキーワードの間で同じ文字の数や割合だけでなく、対

象となるキーワードに関連する他のキーワードの類似性も考慮したマクロ類似度を用いる。これにより、たとえば「伊達公子」と「伊達選手」と「伊達政宗」のように、一致している文字数が等しいキーワードであっても、同一選手を意味する「伊達公子」と「伊達選手」のマクロ類似度は大きく、別人である「伊達公子」と「伊達政宗」はマクロ類似度が小さくなる。この結果、別表現されているキーワードであっても同じ意味を表しているキーワードは類似しているとし、文字だけをみれば類似しているキーワードであっても別の話題でべつの事柄を表しているものは、あまり類似していないとすることができる。これらは、マクロ類似度といった数値の大きさに判断され、この判断に基づいて加算された頻度を用いて要約結果を得ることができる。

【0264】第8の発明にかかる実施の形態によれば、文字放送といった複数の放送局からデータが提供され、データを作成した作成者が異なる場合において別表現されたキーワードの処理が可能である。文字放送では、たとえばニュースにおける会議名、人名、会社名など、同じようなキーワードで、全く関係のない話題で扱われていることがある。したがって、マクロ類似度により同じ意味内容で別表現されているキーワードと、同じような表現であるが全く別の意味のキーワードを、それぞれ類似キーワード、類似していないキーワードとして処理する機能が必要があり、マクロ類似度を用いることにより、この機能を実現している。

【0265】第9の発明にかかる実施の形態によれば、要約を表すキーワードとして適切でないものを予め例外キーワード記憶部91に記憶している。この例外キーワード記憶部91に記憶されているキーワードを取り除いてからキーワードを処理し、キーワードの頻度に基づいて要約を行うために、たとえば英語の{a, the}などの冠詞、前置詞など話題を表す上で意味のないキーワードを取り除くことができる。第10の発明にかかる実施の形態によれば、文字放送において、たとえば英語の引用文が提供される場合でも、例外キーワード記憶部91に記憶しているキーワードを取り除いて、要約を行う。この結果、英語における前置詞、冠詞など話題を表す上で重要でないキーワードが要約結果に出ないようにすることができる。

【0266】

【発明の効果】以上述べたところから明らかなように本発明は、従来にくらべてより一層適切なキーワードをデータ内から抽出することが出来るという長所を有する。

【図面の簡単な説明】

【図1】第1の発明にかかる実施の一形態である情報要約方法、および情報要約装置のシステム構成を示すブロック図である。

【図2】同実施の一形態におけるハードウェア構成例である。

【図 3】同実施の一形態における情報要約方法、および情報要約装置の動作手順を示すフローチャートである。

【図 4】同実施の一形態における入力データのデータ例である。

【図 5】同実施の一形態における入力データのデータ例である。

【図 6】同実施の一形態における出力例である。

【図 7】第 2 の発明にかかる実施の一形態である文字放送受信装置のシステム構成を示すブロック図である。

【図 8】同実施の一形態におけるハードウェア構成例である。 10

【図 9】同実施の一形態における文字放送受信装置の動作手順を示すフローチャートである。

【図 10】第 3 の発明にかかる実施の一形態である情報要約方法、および情報要約装置のシステム構成を示すブロック図である。

【図 11】同実施の一形態における情報要約方法、および情報要約装置の動作手順を示すフローチャートである。

【図 12】同実施の一形態における出力例である。

【図 13】第 4 の発明にかかる実施の一形態である文字放送受信装置のシステム構成を示すブロック図である。

【図 14】同実施の一形態における文字放送受信装置の動作手順を示すフローチャートである。

【図 15】キーワードの頻度のみに基づく情報要約の出力例である。

【図 16】第 5 の発明にかかる実施の一形態である情報要約方法、および情報要約装置のシステム構成を示すブロック図である。

【図 17】同実施の一形態における情報要約方法、および情報要約装置の動作手順を示すフローチャートである。 30

【図 18】第 6 の発明にかかる実施の一形態である文字放送受信装置のシステム構成を示すブロック図である。

【図 19】同実施の一形態における文字放送受信装置の動作手順を示すフローチャートである。

【図 20】第 7 の発明にかかる実施の一形態である情報要約方法、および情報要約装置のシステム構成を示すブロック図である。

【図 21】同実施の一形態における情報要約方法、および情報要約装置の動作手順を示すフローチャートである。 40

【図 22】第 8 の発明にかかる実施の一形態である文字放送受信装置のシステム構成を示すブロック図である。

【図 23】同実施の一形態における文字放送受信装置の動作手順を示すフローチャートである。

【図 24】第 9 の発明にかかる実施の一形態である情報要約方法、および情報要約装置のシステム構成を示すブロック図である。

【図 25】同実施の一形態におけるハードウェア構成例 50

である。

【図 26】同実施の一形態における情報要約方法、および情報要約装置の動作手順を示すフローチャートの一部分である。

【図 27】第 10 の発明にかかる実施の一形態である文字放送受信装置のシステム構成を示すブロック図である。

【図 28】同実施の一形態におけるハードウェア構成例である。

【図 29】第 9 の発明の第 2 の実施の一形態、第 3 の実施の一形態、第 4 の実施の一形態のシステム構成の一部である。

【図 30】第 10 の発明の第 2 の実施の一形態、第 3 の実施の一形態、第 4 の実施の一形態のシステム構成の一部である。

【図 31】データ検索機能を備えた情報要約方法、および情報要約装置のシステム構成の一部である。

【図 32】データ検索機能を備えた文字放送受信装置のシステム構成の一部である。

【図 33】データの到着を知らせる機能を備えた情報要約方法、および情報要約装置のシステム構成の一部である。

【図 34】データの到着を知らせる機能を備えた情報要約方法、および情報要約装置の動作手順の一部を示すフローチャートである。

【図 35】データの到着を知らせる機能を備えた情報要約方法、および情報要約装置の他の動作手順の一部を示すフローチャートである。

【図 36】データの到着を知らせる機能を備えた情報要約方法、および情報要約装置においてデータの到着を知らせる出力例を表した図である。

【図 37】英語で記述された入力データを表すデータ例である。

【図 38】英語で記述された入力データを処理した場合の出力例である。

【図 39】英語で記述された入力データをキーワードどうしの関連付けを考慮して処理した場合の出力例である。

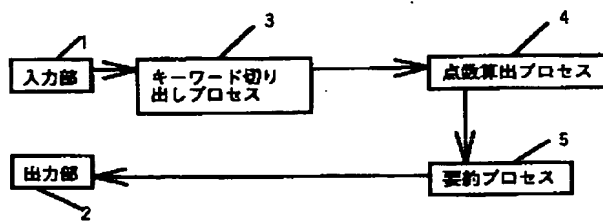
【符号の説明】

- | | |
|----|-----------------------|
| 1 | 入力部 |
| 2 | 出力部 |
| 3 | キーワード切り出し部（キーワード抽出手段） |
| 4 | 点数算出部（重み付け手段） |
| 5 | 要約部（キーワード選択手段） |
| 11 | VRAM |
| 12 | 主記憶装置 |
| 13 | 外部記憶装置 |
| 14 | CPU |
| 21 | 文字放送受信部 |
| 22 | チャンネル記憶部 |

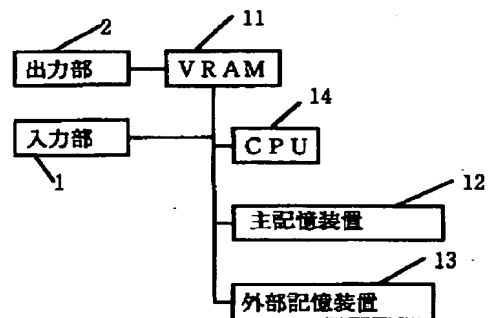
- 2 3 文字放送キーワード切り出し部（キーワード抽出手段）
 2 4 文字放送点数算出部（重み付け手段）
 2 5 文字放送要約部（キーワード選択手段）
 2 6 表示部
 3 1 キーワード関連付部
 3 2 関連付点数算出部（重み付け手段）
 3 3 関連付要約部（選択手段）
 4 1 文字放送キーワード関連付け部（キーワード関連付け手段）
 4 2 文字放送関連付け点数算出部（重み付け手段）
 4 3 文字放送関連付要約部（選択手段）
 5 1 類似度付点数算出部（類似性算出手段、重み付け手段）
 6 1 文字放送類似度付点数算出部（類似性算出手段、

- 重み付け手段）
 7 1 類似度付関連付点数算出部（類似性算出手段、重み付け手段）
 8 1 文字放送類似度付関連付け点数算出部（類似性算出手段、重み付け手段）
 9 1 例外キーワード記憶部
 9 2 例外付キーワード切り出し部
 1 0 1 例外付文字放送キーワード切り出し部
 1 1 1 データ指定手段（キーワード指定手段）
 1 1 2 データ記憶部（作業領域）
 1 1 3 検索部
 1 1 4 データ提示部（出力手段）
 1 2 1 キーワード発見部
 1 2 2 連絡部（出力手段）

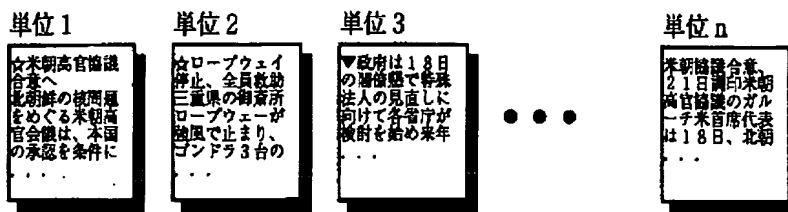
【図 1】



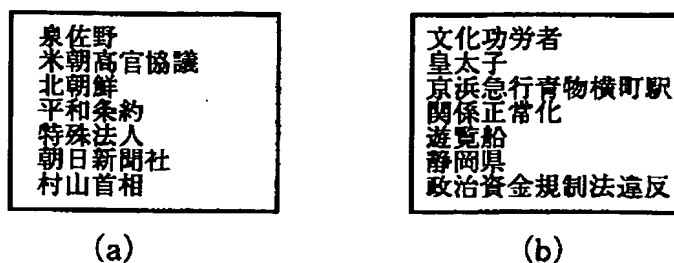
【図 2】



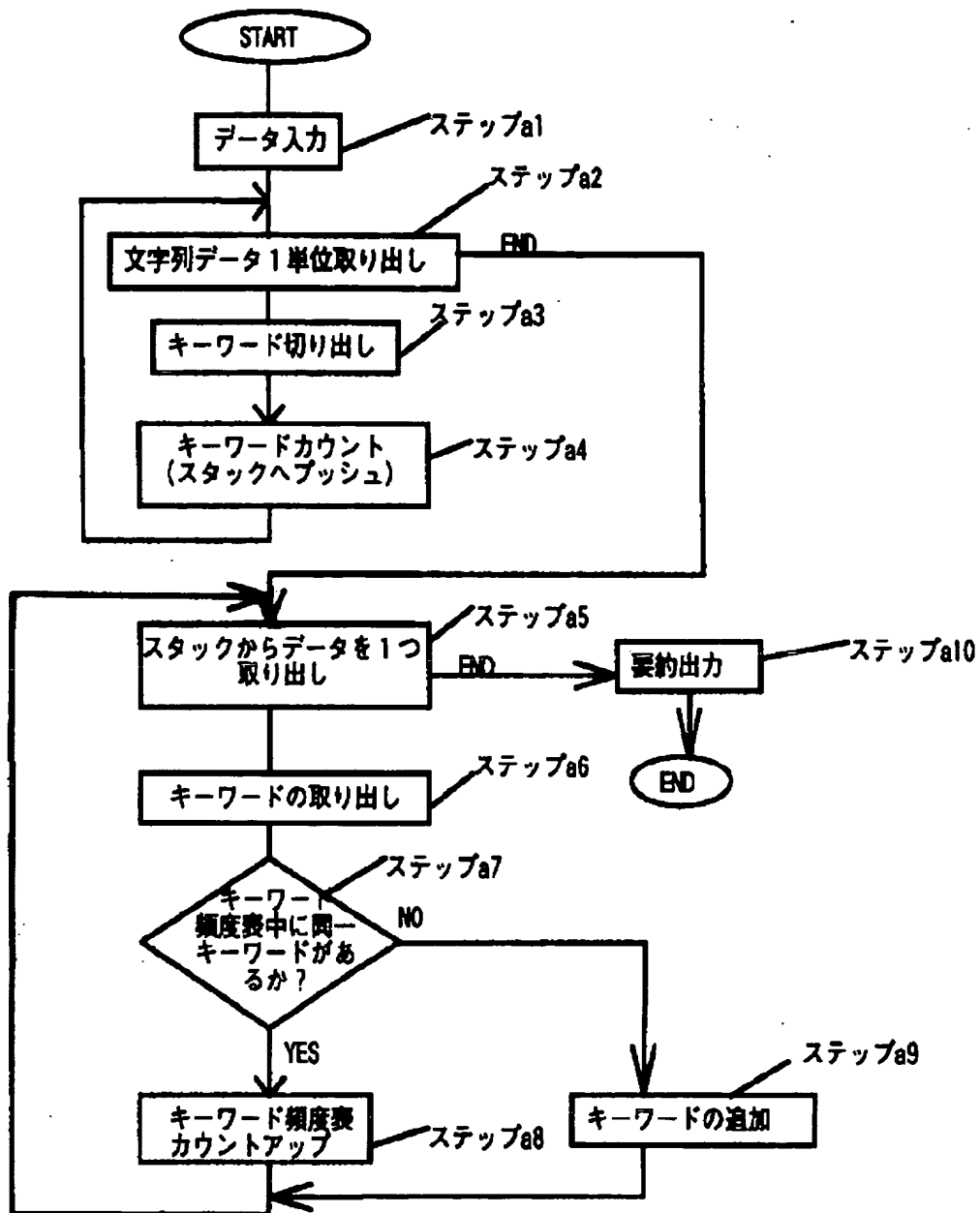
【図 4】



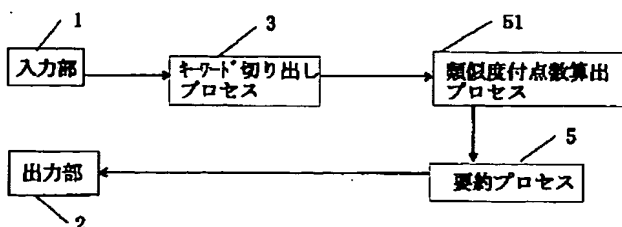
【図 15】



【図 3】



【図 16】



【図5】

単位 1

☆泉佐野でヘリ墜落
大阪府警によると、泉佐野市で朝日新聞社取材用のヘリコプターが墜落、3人が死亡した。

☆新聞大会開く
第47回新聞大会が、京都市で開幕、新聞販売の正常化を図る特別宣言が採択された。

☆電動茶せんを開発
熟練と細かい手の動きが必要な抹茶たてを手軽にできるようにし、抹茶をもっと広めようと、京都府茶協同組合が電動茶せんを開発した。

☆米朝高官協議合意へ
北朝鮮の核問題をめぐる米朝高官協議は、本国の承認を条件に合意に達し、承認されれば21日、ジュネーブで調印式が行われることになった。

単位 2

税制改革審議入り／授与を断念
◆衆院は税制改革関連4法案が審議入り。村山首相は「当面考えられる最前の改革」で、公約違反を否定。
◆与謝野文相は、大江健三郎氏への文化勲章授与を断念。「戦後民主主義は自体も自由」。再考は促さず。

特殊法人見直し／叙勲廃止論
◆特殊法人見直しの最終報告は、来年2月10日までに。政府はこの報告を受け、今年度中に見直し案策定。
◆井手厚相が記者会見で叙勲廃止論。「勲一等とか、国が人の一生を番号付けするのはいかなるものか」。

核問題で合意／平和条約に仮調印
◆米朝高官協議で「核」包括解決に合意。使用済み燃料棒は最終的に移転し廃棄。特別査定は先送り。
◆イスラエルとヨルダンが、両国関係を正常化する平和条約に仮調印。77ヶ諸国の中で、79年のエジプトに次ぎ2カ国目。

単位 3

▼朝鮮民主主義人民共和国の核疑惑を巡る米朝高官協議で「ルーチ米代表は17日夜、原則合意に達し21日に双方がジュネーブで調印予定」と述べた。
▼北朝鮮が将来のIAEA査察受け入れを条件に米が軽水炉支援を先行実施するなどが内容とみられる。

▼18日午前大阪・泉佐野市で朝日新聞社のヘリが墜落し3人が死亡した事故は毎日新聞社のヘリと接触した直後に墜落したことが分かった。
▼18日の参院予算委で村山首相は政府の電力供給計画に沿った原発の増設を認める考えを示した。

▼富士写真フイルム専務殺害事件で指名手配されていた山口組系の暴力団員が18日午後警察に出頭して犯行を認め殺人容疑で逮捕された。
▼逮捕されたのは岡本大介容疑者(25)で「沖野容疑者に頼まれてやった。後で5万円もらった」と供述。

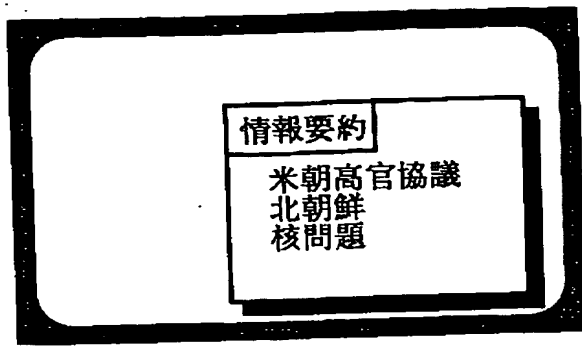
単位 4

米朝協議合意、21日調印
米朝高官協議のルーチ米首席代表は18日、北朝鮮の核問題で合意、21日に調印すると発表。
全面的検査の受け入れと核拡防条約への完全復帰、軽水炉支援など包括的合意に達した模様。

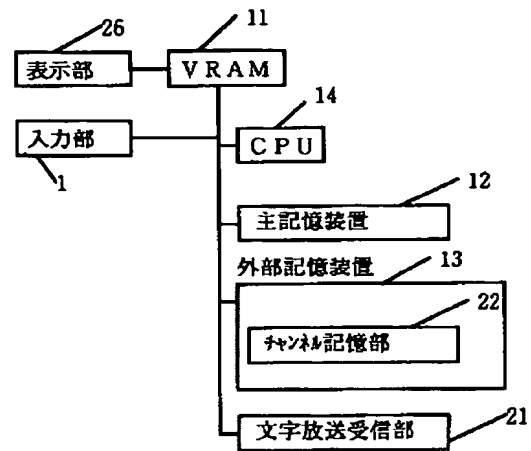
日韓外相、合意内容を評価
韓国外相は、米朝合意に核開発凍結の手順と南北対話の必要性を盛り込んでいることを評価、南北対話の年内再開に期待を表明。
河野外相も「指示できる内容」と対話による解決を歓迎。

イラク議長も指示声明
イラク国会議長は17日、政府が表明したクウェートの主権、国境線承認の方針を支持する声明。しかし、正式な承認決定発表までは対イラク経済制裁解除を国際社会が受け入れる見込みはない。

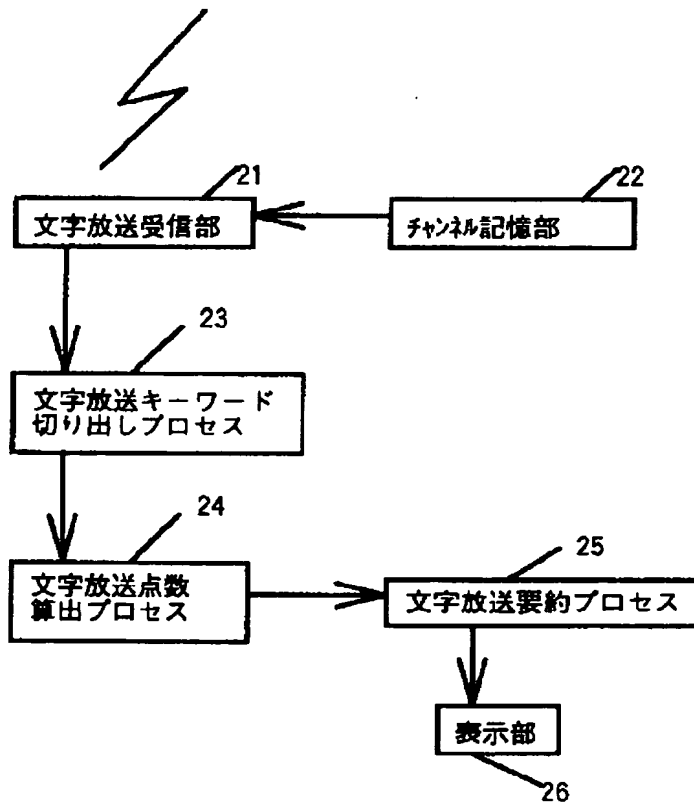
【図6】



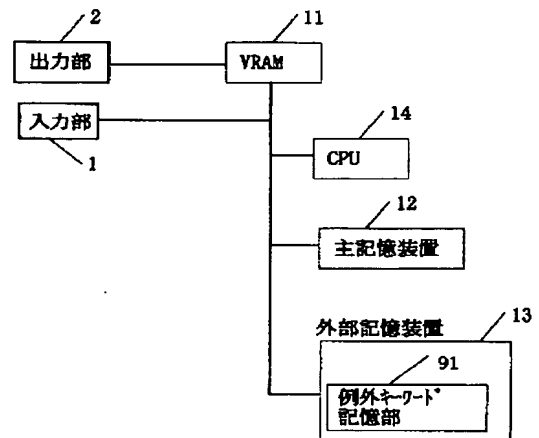
【図8】



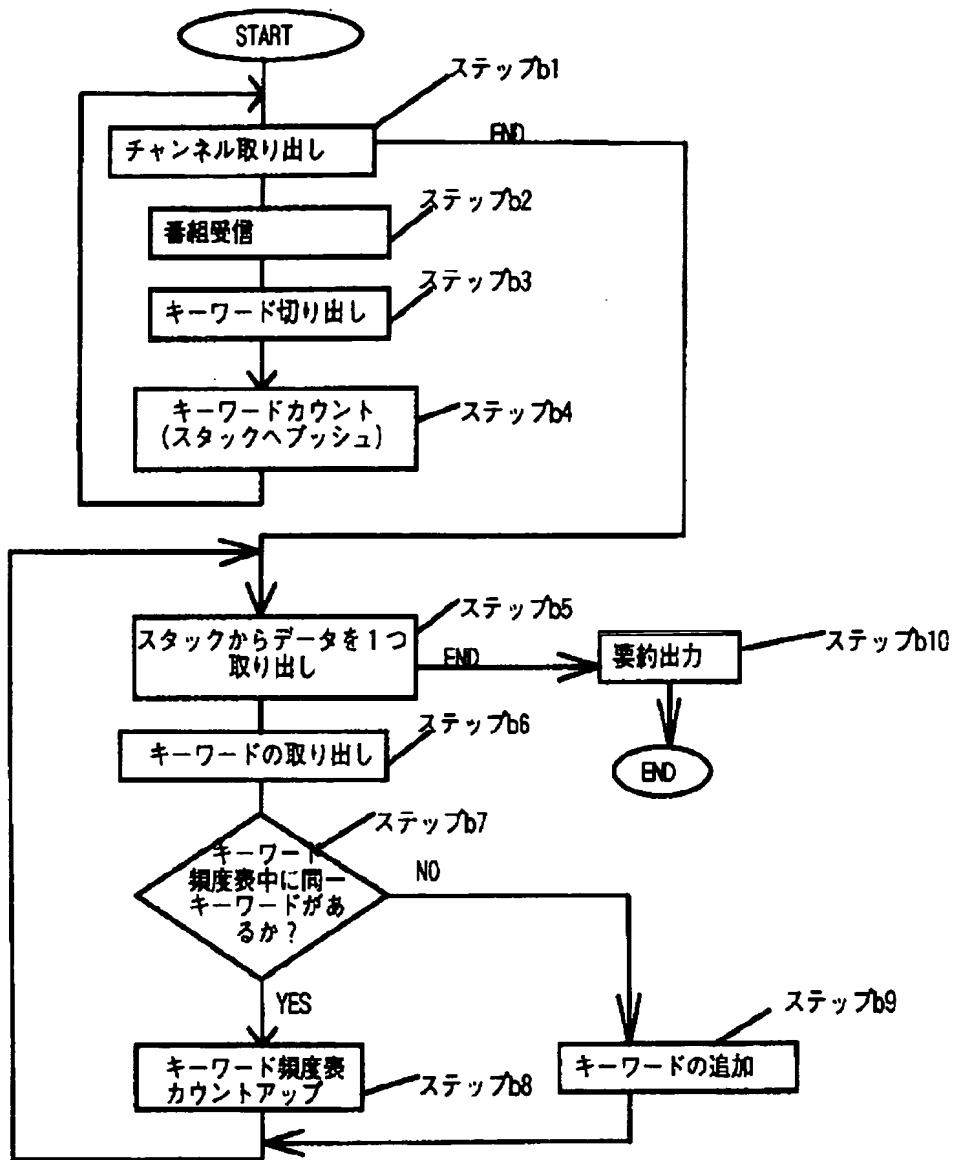
【図7】



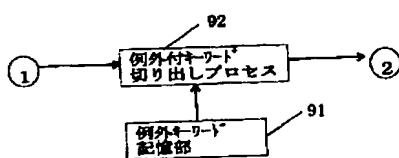
【図25】



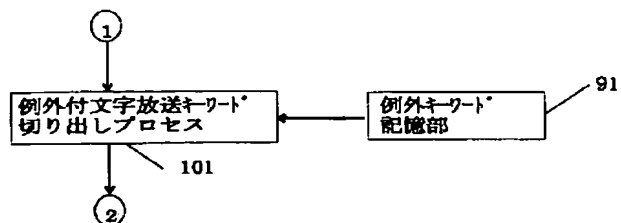
【図9】



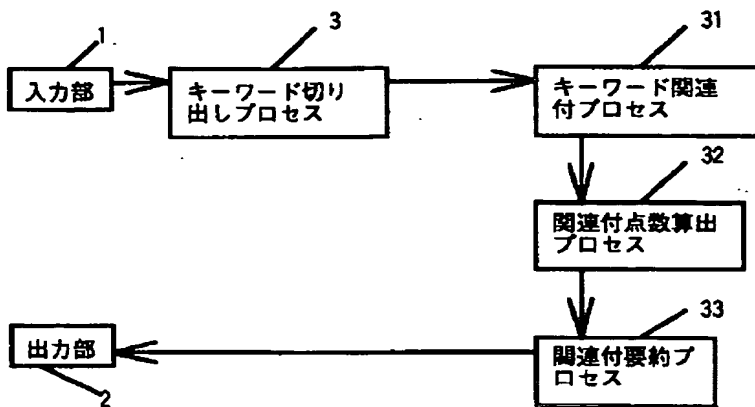
【図29】



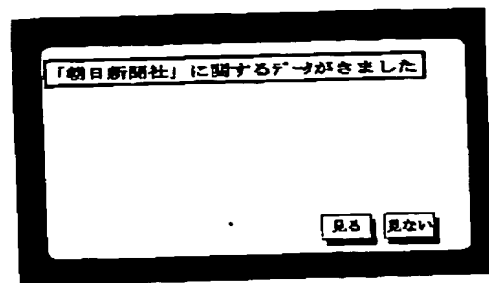
【図30】



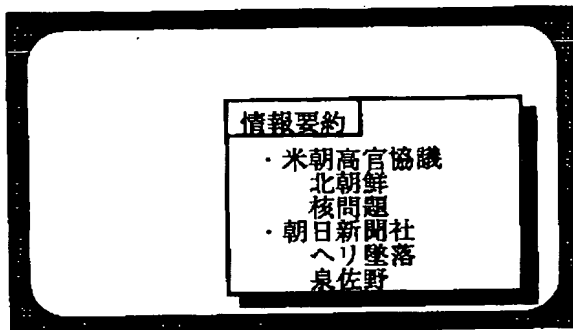
【図10】



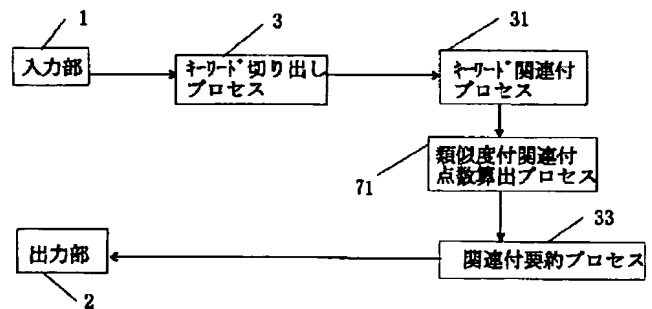
【図36】



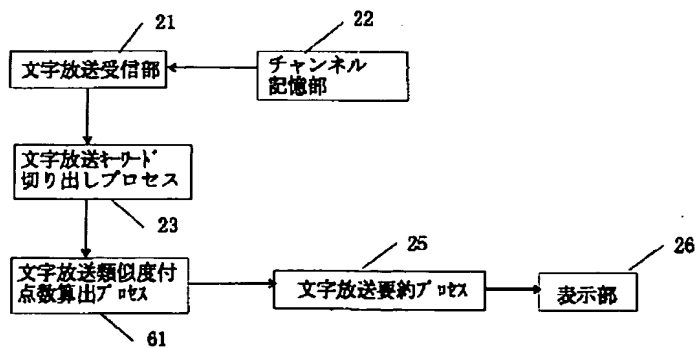
【図12】



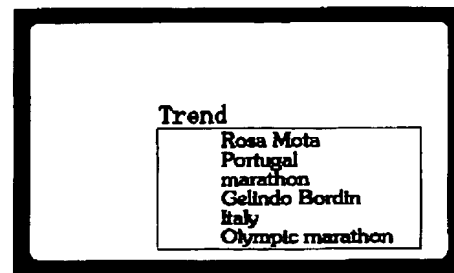
【図20】



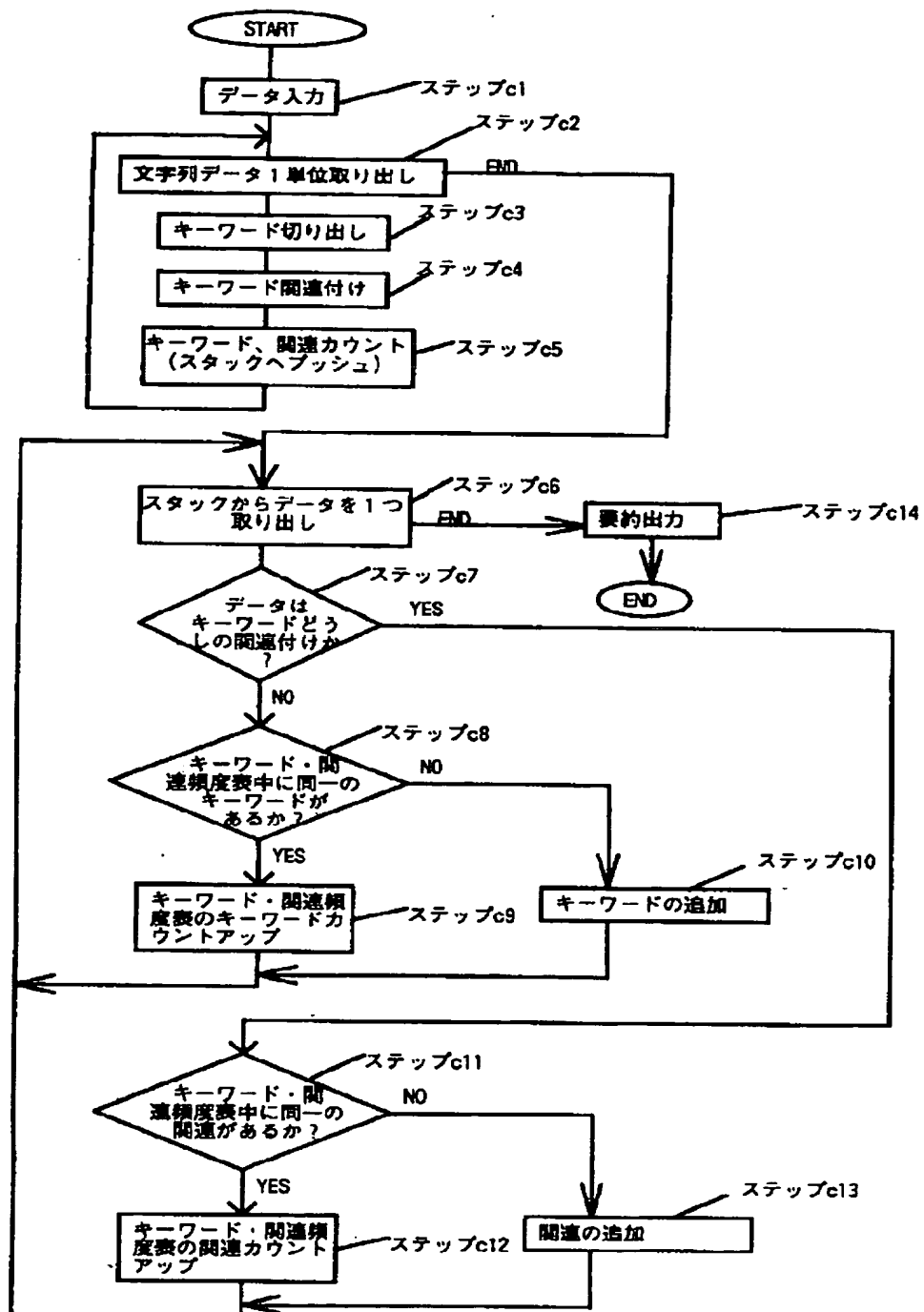
【図18】



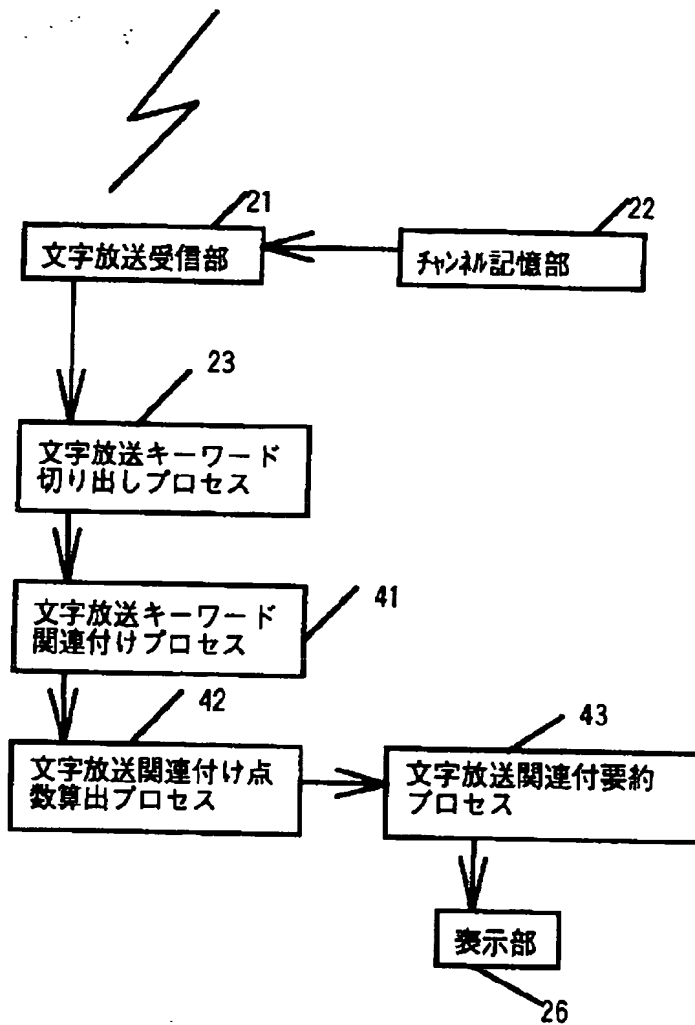
【図38】



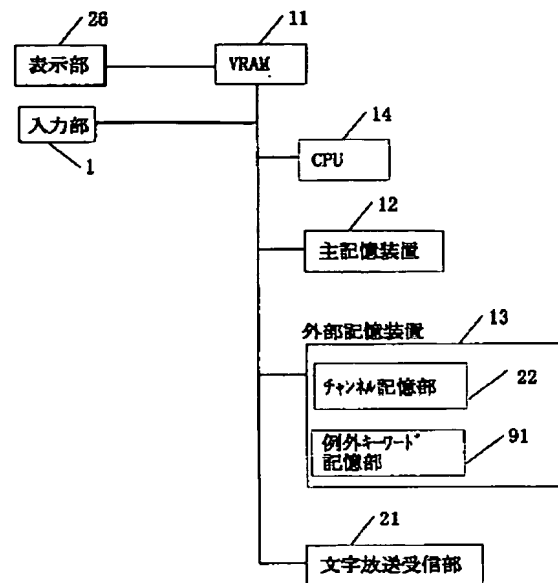
【図11】



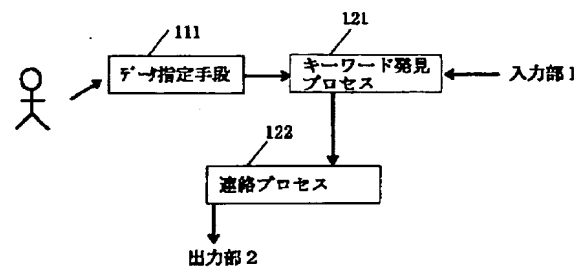
【図 1 3】



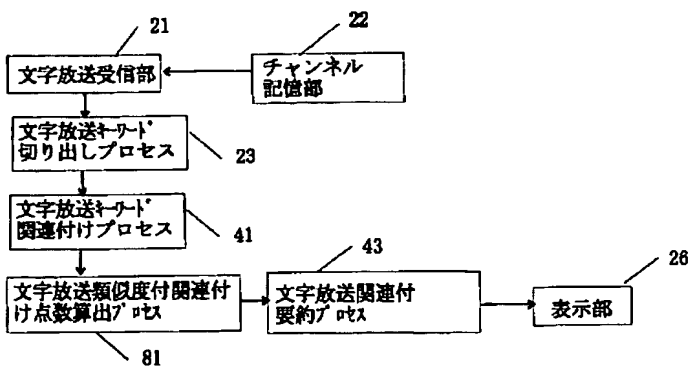
【図 2 8】



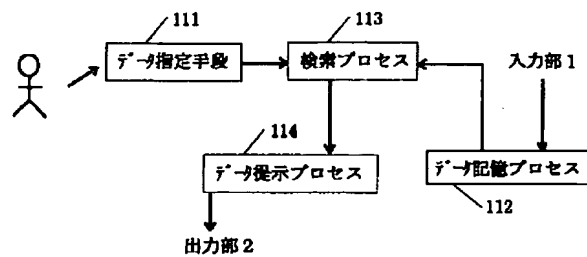
【図 3 3】



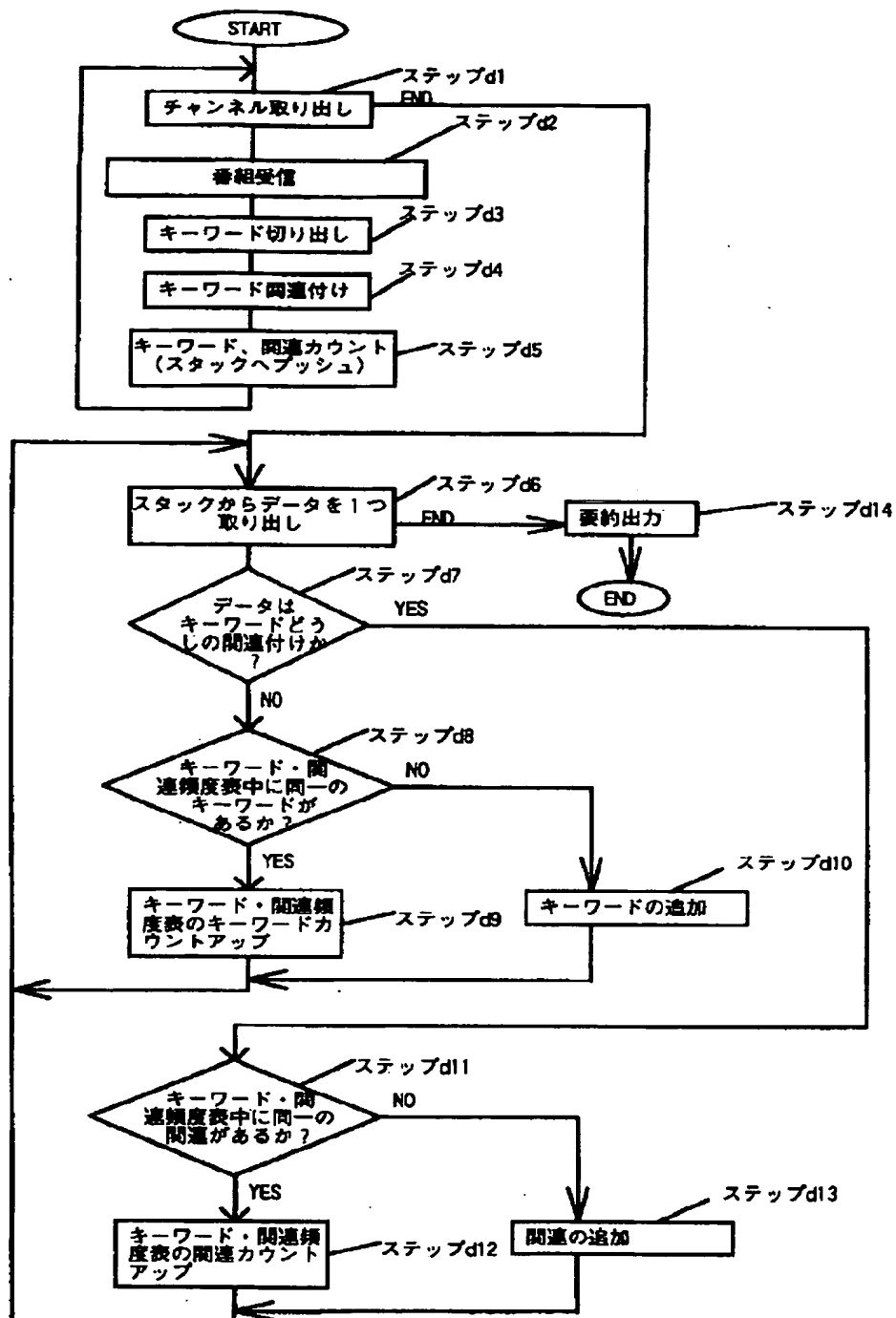
【図 2 2】



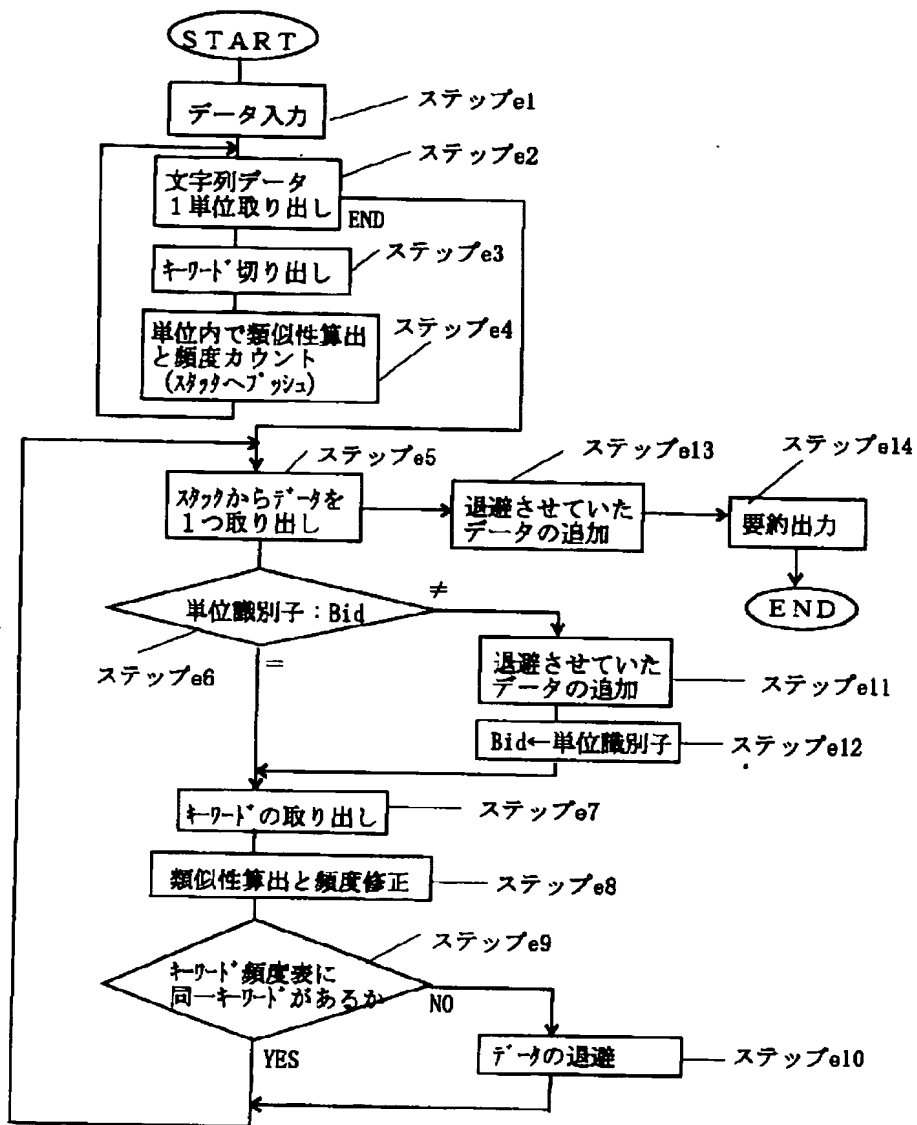
【図 3 1】



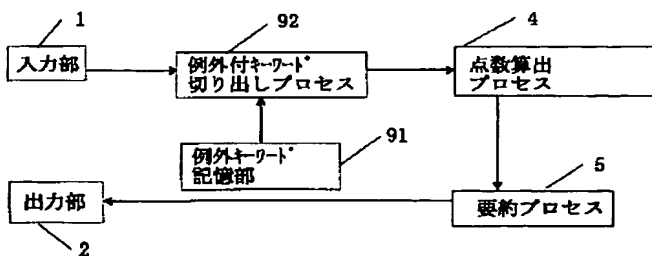
【図14】



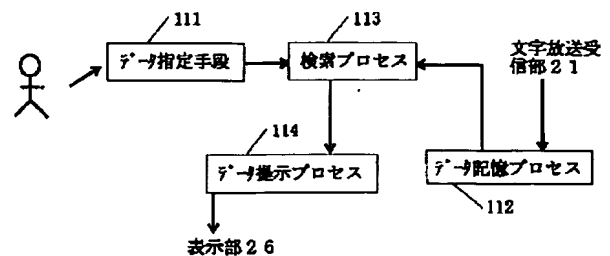
【図17】



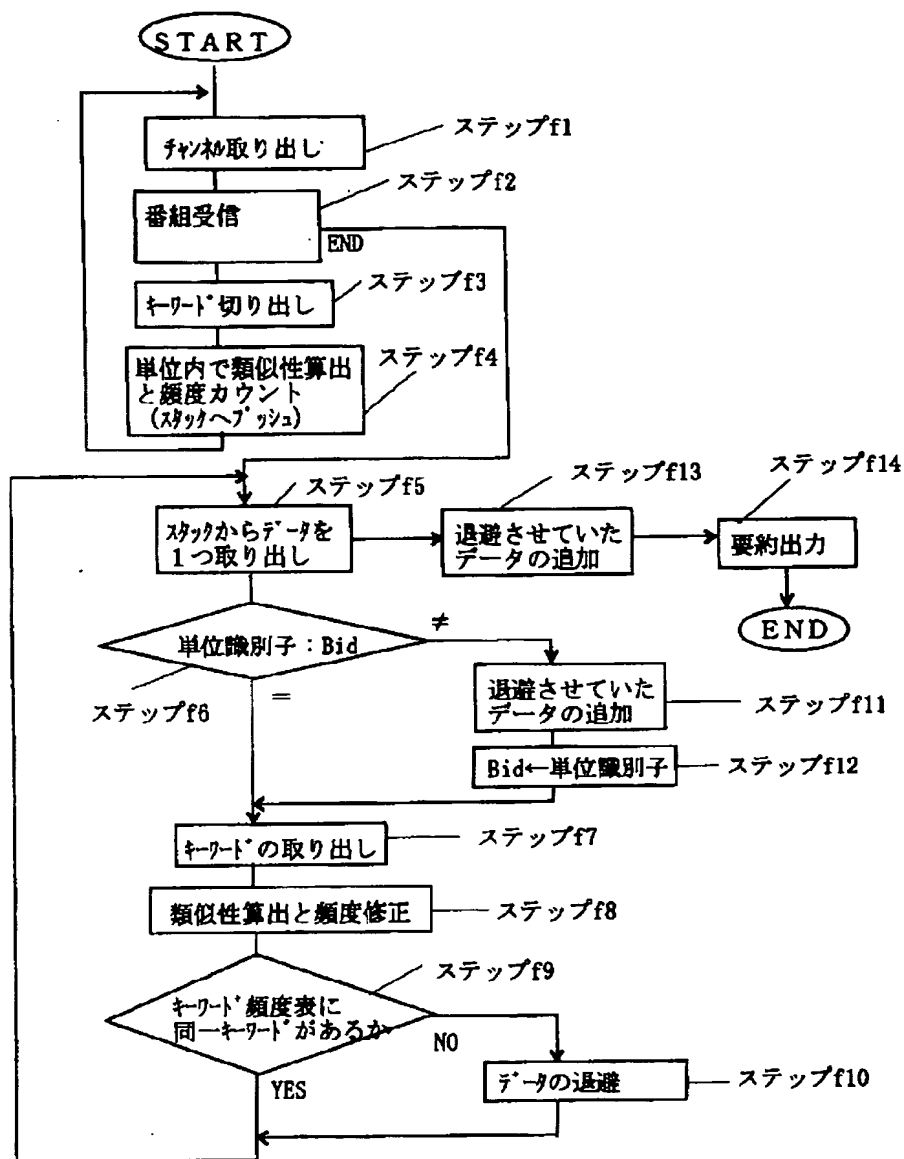
【図24】



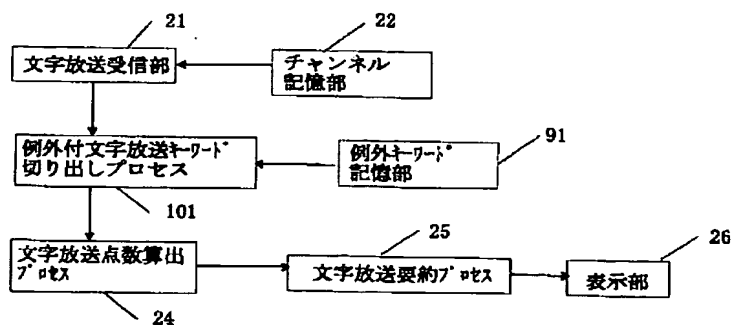
【図32】



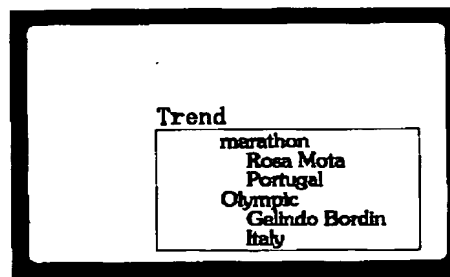
【図19】



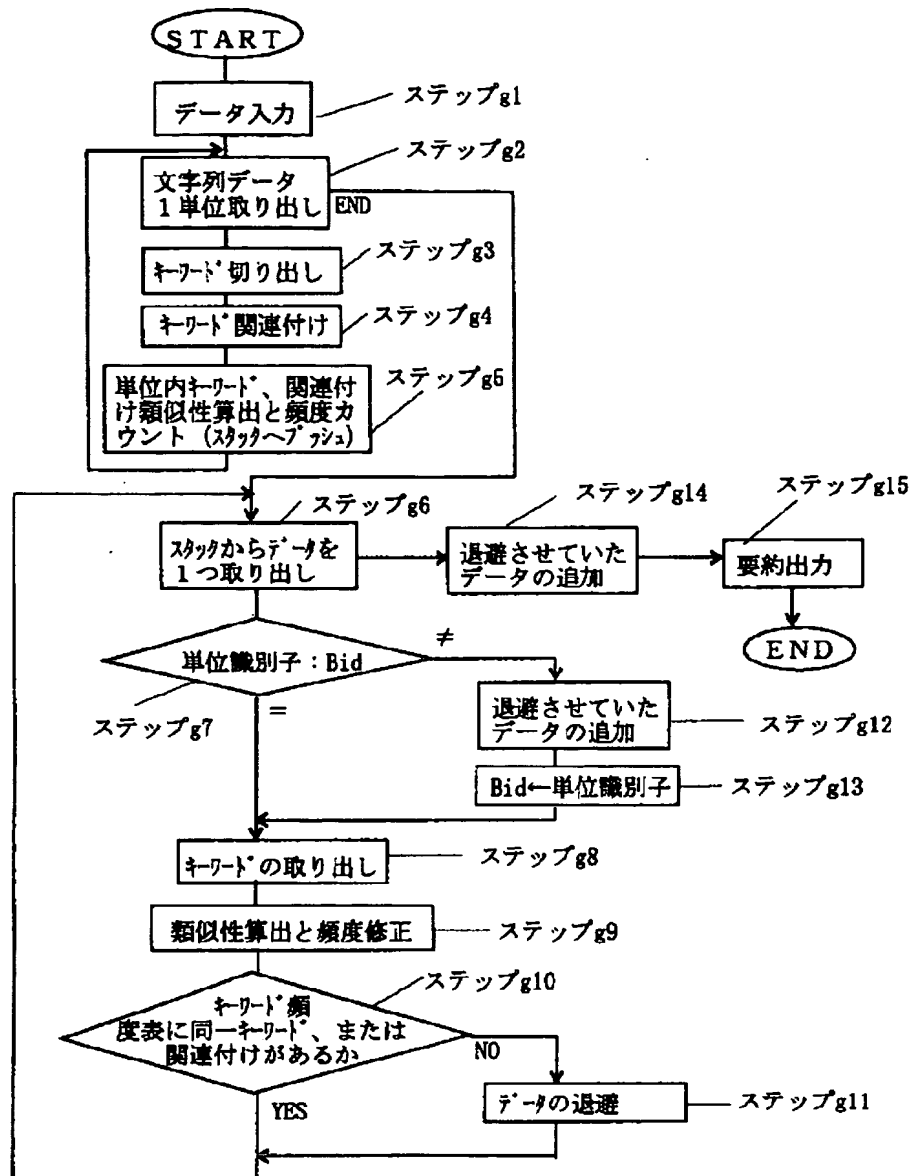
【図27】



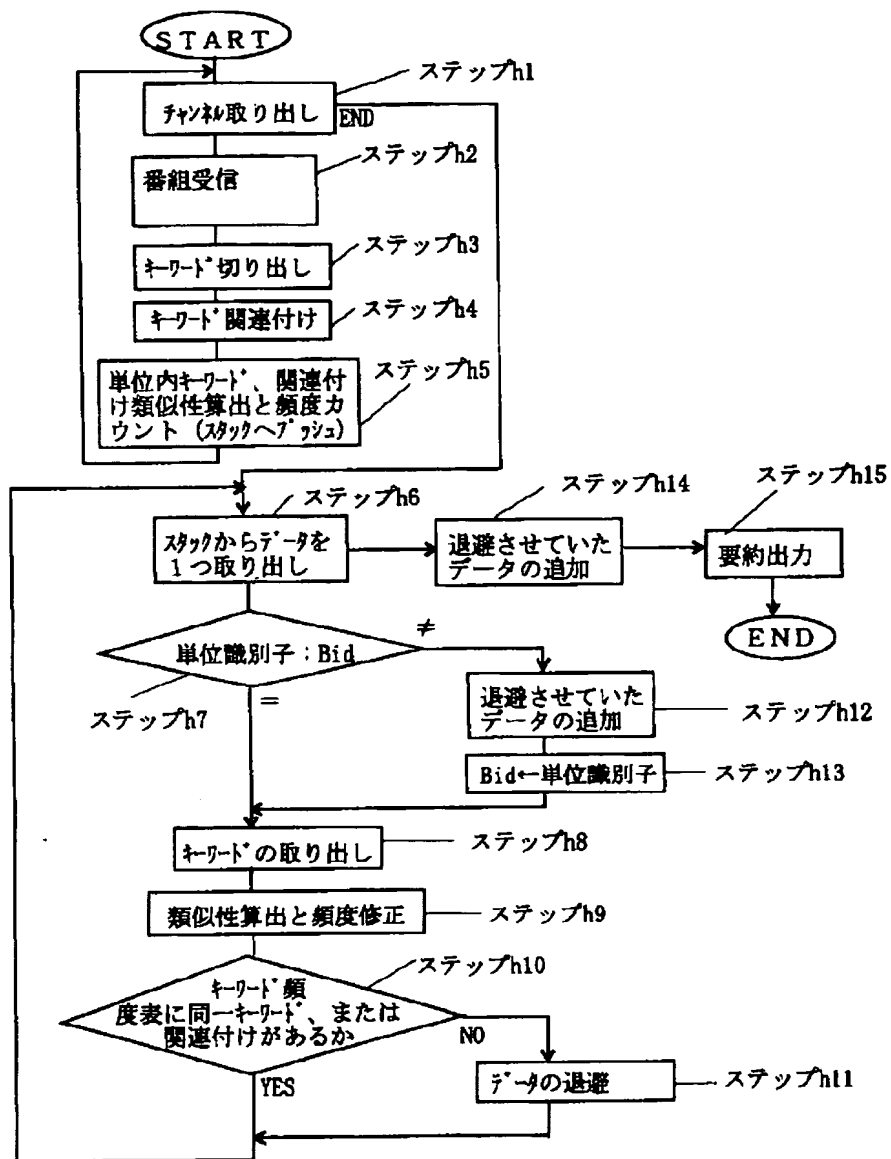
【図39】



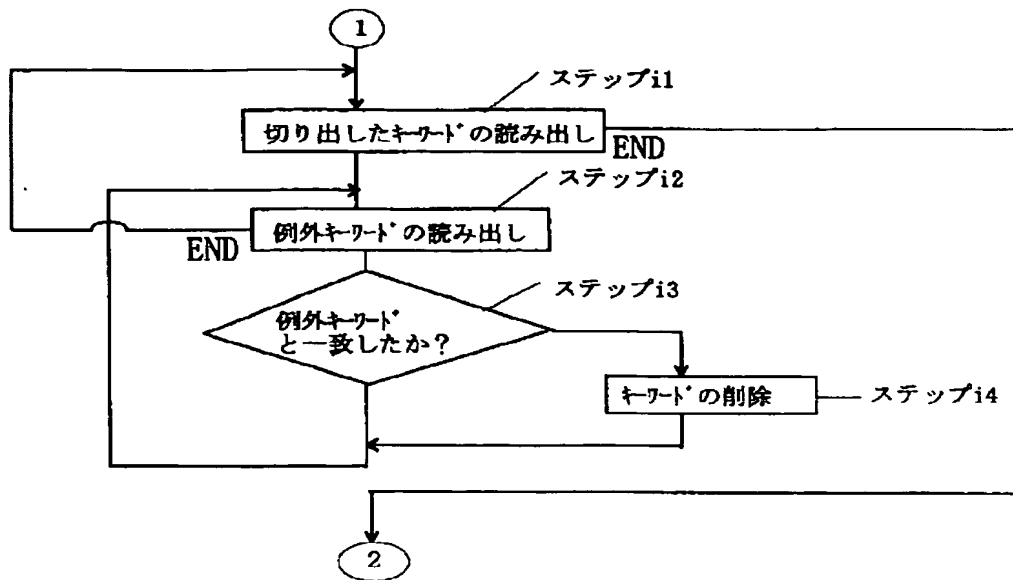
【図21】



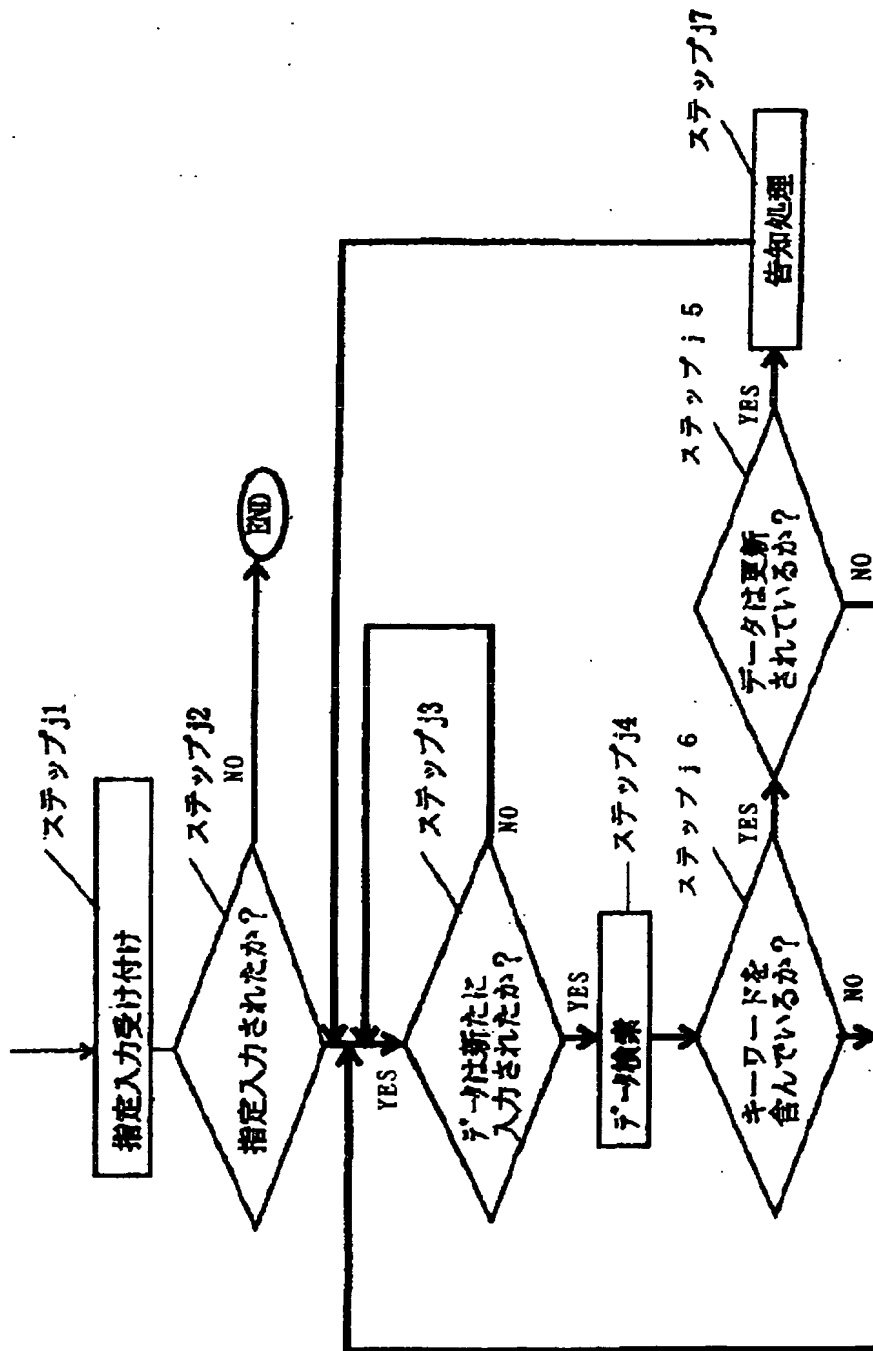
【図23】



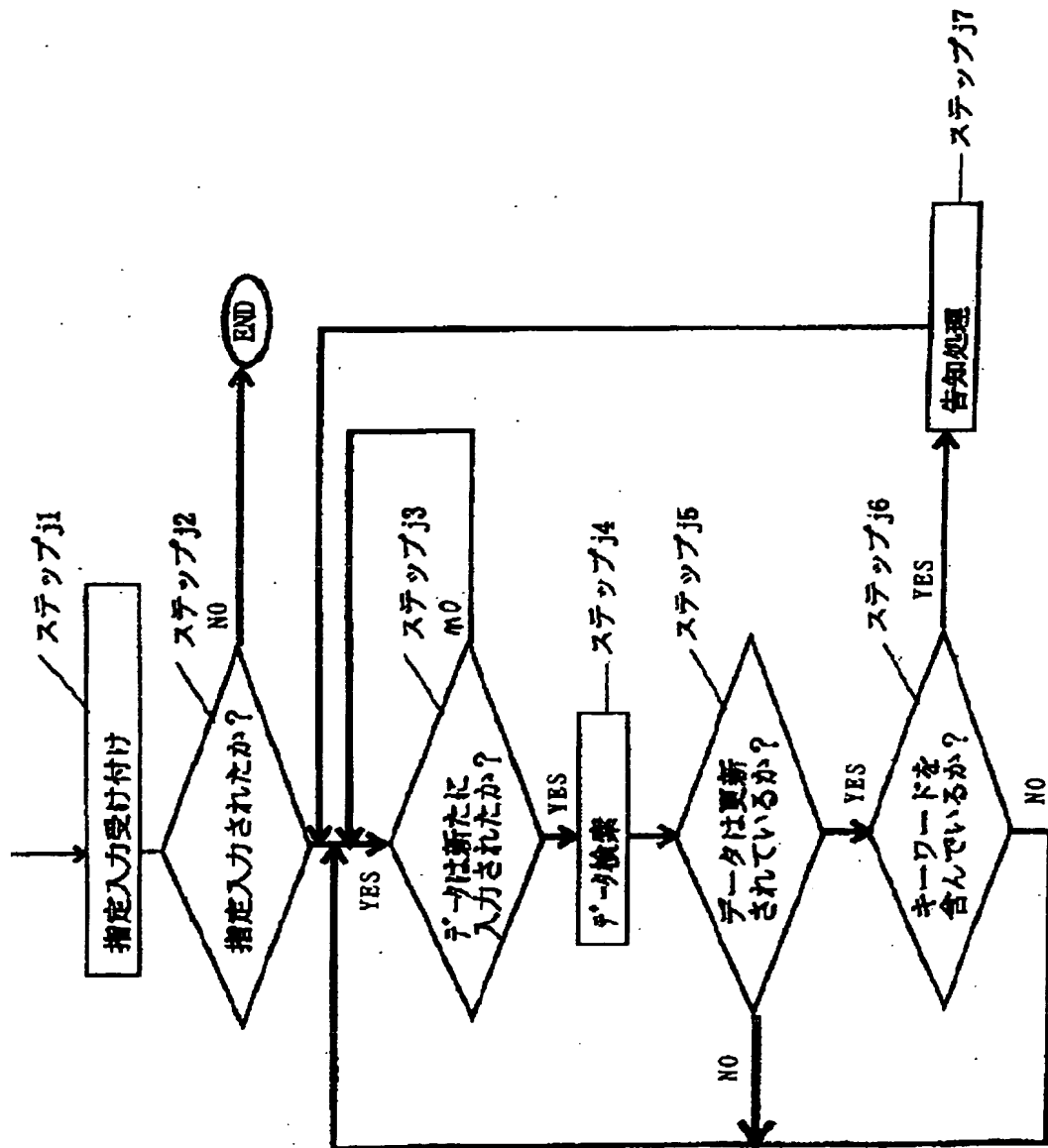
—274—



【図34】



【図35】



【図37】

単位 1

☆1984 U.S. Olympic champion
Joan Benoit Samuelson featured
in NYC Marathon.

☆Rosa Mota of Portugal wins
Olympic marathon.

単位 3

◆A look at the homeless problem
in South Korea, made worse by
preparations for the Olympic
Games.

◆Portugal's Rosa Mota wins the
women's Olympic marathon;
other track and field results
included.

単位 2

▼U.S. runner Carlos Talbott
wins gold medal in marathon
as Intl. Paralympic Games end.

▼Gelindo Bordin of Italy wins
Olympic marathon.

単位 4

Italy's Gelindo Bordin
overtakes Ahmed Salah of
Djibouti during final mile,
wins Olympic marathon.

Portugal's Rosa Mota wins
marathon